

# Modelling Spatial Compositional Data: Reconstructions of past land cover and uncertainties

Behnaz Pirzamanbein \* <sup>1,2</sup>, Johan Lindström<sup>1</sup>, Anneli Poska<sup>3,4</sup> and  
Marie-José Gaillard<sup>5</sup>

<sup>1</sup>Centre for Mathematical Sciences, Lund University, Sweden

<sup>2</sup>Centre for Environmental and Climate Research, Lund University, Sweden

<sup>3</sup>Department of Physical Geography and Ecosystems Analysis, Lund University,  
Sweden

<sup>4</sup>Institute of Geology, Tallinn University of Technology, Estonia

<sup>5</sup>Department of Biology and Environmental Sciences, Linnaeus University, Sweden

## Abstract

In this paper, we construct a hierarchical model for spatial compositional data, which is used to reconstruct past land-cover compositions (in terms of coniferous forest, broadleaved forest, and unforested/open land) for five time periods during the past 6 000 years over Europe. The model consists of a Gaussian Markov Random Field (GMRF) with Dirichlet observations. A block updated Markov chain Monte Carlo (MCMC), including an adaptive Metropolis adjusted Langevin step, is used to estimate model parameters. The sparse precision matrix in the GMRF provides computational advantages leading to a fast MCMC algorithm. Reconstructions are obtained by combining pollen-based estimates of vegetation cover at a limited number of locations with scenarios of past deforestation and output from a dynamic vegetation model. To evaluate uncertainties in the predictions a novel way of constructing joint confidence regions for the entire composition at each prediction location is proposed. The hierarchical model's ability to reconstruct past land cover is evaluated through cross validation for all time periods, and by comparing reconstructions for the recent past to a present day European forest map. The evaluation results are promising and the model is able to capture known structures in past land-cover compositions.

*Keywords:* Gaussian Markov Random Field, Dirichlet Observation, Adaptive Metropolis adjusted Langevin, Pollen records, Confidence regions.

---

\*This research is part of two Swedish strategic research areas: Biodiversity and Ecosystems in a Changing Climate (BECC) and Modelling the Regional and Global Earth system (MERGE). The paper is also a contribution to PAGES LandCover6k. Poska received financial support from FORMAS LUSIT project and ETF project 9031. Gaillard received financial support from the Faculty of Health and Life Sciences, Linnaeus University, Kalmar, Sweden.

# 1 Introduction

Modelling the spatial distribution in species composition and the relative abundances of different species is a common problem in environmental studies. (Aitchison, 1986; Paciorek and McLachlan, 2009; Billheimer et al., 2001; Tjelmeland and Lund, 2003; Pirzamanbein et al., 2014). In this paper we develop a statistical model for spatial compositional data and a way of assessing the uncertainties in the resulting compositional reconstructions at unobserved locations. The model is used to reconstruct past land-cover composition (LCC) over Europe from local pollen-based estimates of vegetation cover.

## 1.1 Spatial Interpolation of Compositional Data

A common approach to modelling compositional data is Gaussian modelling of log-ratio transformed data (Aitchison, 1986), where the spatial structure can be captured using Gaussian fields (Billheimer et al., 2001; Tjelmeland and Lund, 2003; Pirzamanbein et al., 2014). However, modelling transformed compositions as Gaussian might understate the uncertainty in the data, especially in cases of over-dispersion (Paciorek and McLachlan, 2009).

To capture the variability in our observations, we propose a Bayesian hierarchical model (BHM, described in Sec. 2) where the compositional data are seen as Dirichlet observations of an underlying latent field of probabilities. Including known location specific weights in the Dirichlet distribution allows us to account for known relative differences in the uncertainties at different observations locations. The field of compositional probabilities is in turn modelled using a transformed Gaussian Markov Random Field (GMRF) (Rue and Held, 2004; Lindgren et al., 2011). The sparsity in the precision matrix of the GMRF allows us to compute the Hessian for the entire latent field, allowing for fast estimation (see Sec. 3) using a Metropolis Adjusted Langevin algorithm (MALA) (Girolami and Calderhead, 2011).

To describe the uncertainties in the compositional reconstructions we propose a novel way of computing joint confidence and prediction regions for compositional data (Sec. 4). The method accounts for the interdependence among the components of the compositional data and allows us to illustrate the joint uncertainty in the composition at each prediction location.

The model and parameter estimation procedure are investigated using a simulation study (Sec. 5.1), before being applied to pollen-based LCC data.

## 1.2 Climate Studies and Past Land Cover

For climate modelling studies the LCC is commonly divided into three land cover types (LCTs): coniferous forest, broadleaved forest, and unforested/open land. The spatial distribution of these LCTs play an important role in the climate system (Claussen et al., 2001). Accurate, spatially continuous, descriptions of past LCTs are necessary to assess past land cover-climate interactions (Brovkin et al., 2006) and the impact of anthropogenic land-cover changes on climate (Strandberg et al., 2014; Gaillard et al., 2010, 2015; Pirzamanbein et al., 2014).

Historic maps and surveys of past land cover have limited temporal coverage (rarely more than the past 300 to 500 years) and is often spatially fragmented due to a lack of transnational databases. Land-cover in climate models is currently

implemented using a combination of dynamic vegetation models (DVMs) (e.g. LPJ-GUESS Smith et al., 2001) and scenarios of anthropogenic land-cover changes (e.g. Kaplan et al., 2009; Klein Goldewijk et al., 2011; Pongratz et al., 2008). Here, the DVMs provide a climate-induced, potential vegetation, which is modified by the anthropogenic scenarios to account for human activities (mainly deforestation).

Land-cover reconstruction from fossil pollen records is an alternative, to the DVM simulations and anthropogenic scenarios, that may provide more realistic descriptions of past land cover for climate modelling studies (Gaillard et al., 2010; Trondman et al., 2015). Given pollen records extracted from lakes and bogs, pollen-based estimates of vegetation cover are obtained using a model (here the REVEALS model of Sugita, 2007a,b). The model provides estimates of pollen-based land-cover compositions (PbLCC) for a limited area (ca. 100 km x 100 km) around each lake or bog. For use in climate modelling these PbLCC estimates need to be interpolated into continuous maps of past LCC at sub-continental to global scales (Pirzamanbein et al., 2014; Paciorek and McLachlan, 2009).

The PbLCC data used in this paper are available for five time periods during the past 6 000 years, and the proposed BHM and estimation procedure is used to interpolate the PbLCC data for each time period. The results are validated using present-time forest maps and cross-validation (Sec. 5.2.3). The model shows good predictive power, capturing known structures and historical changes in LCC.

The paper ends with some brief conclusion in Sec. 6.

## 2 Model

To model the spatial structure in the compositional data we propose a BHM, where the observed compositions at each location are modelled as draws from a Dirichlet distribution. The Dirichlet is parametrized using a scale (or concentration) parameter and a vector of probabilities. The spatial dependence in these compositional probabilities is modelled using a transformed GMRF. Details regarding the observational model are given in Sec. 2.1, and Sec. 2.2 describes the latent field.

### 2.1 Dirichlet Distribution and Link Function

Compositional data are discussed in detail by Aitchison (1986), here a brief overview is given. Let  $\mathbf{y}_s = (y_{s,1}, y_{s,2}, \dots, y_{s,D})$  be the  $D$ -compositional data at location  $u_s \in \mathbb{R}^2$ ,  $s = 1, \dots, N_o$ , the restrictions for compositional data imply that:  $y_{s,k} \in (0, 1)$  and  $\sum_{k=1}^D y_{s,k} = 1$ . Conditional on the transformed underlying field,  $\mathbf{z} = f(\boldsymbol{\eta})$ , we assume that the data,  $\mathbf{Y} = \{\mathbf{y}_s\}_{s=1}^{N_o}$ , are independent draws from a multivariate Dirichlet distribution,

$$\mathbb{P}(\mathbf{Y}|\boldsymbol{\alpha}, \mathbf{w}, \mathbf{z}) = \prod_{s=1}^{N_o} \left( \frac{\Gamma(\boldsymbol{\alpha}\mathbf{w}_s)}{\prod_{k=1}^D \Gamma(\boldsymbol{\alpha}\mathbf{w}_s z_{s,k})} \prod_{k=1}^D y_{s,k}^{\boldsymbol{\alpha}\mathbf{w}_s z_{s,k} - 1} \right), \quad \boldsymbol{\alpha}, \mathbf{w} > 0, \quad (1)$$

and

$$z_{s,k} \in (0, 1), \quad \sum_{k=1}^D z_{s,k} = 1$$

where  $\alpha$  is a Dirichlet scale parameter. The weight vector  $\mathbf{w}$  describes the known relative uncertainty between different observation locations, if no difference in uncertainty exists then  $w_s \equiv 1$ .

The link function,  $f$ , between  $\mathbf{z}$  and  $\boldsymbol{\eta}$  can be any function from  $\mathbb{R}^{d \times N_o}$  to  $(0, 1)^{D \times N_o}$  such that:

$$f(\eta_1, \dots, \eta_d) = (Z_1, \dots, Z_d, Z_D), \quad \sum_{k=1}^D Z_k = 1, \quad \text{and} \quad d = D - 1. \quad (2)$$

Here  $Z_k$  is a  $N_o \times 1$  column vector containing the  $k^{\text{th}}$  component of the D-compositional data and  $\eta_k$  is a column vector with the  $k^{\text{th}}$  latent field, i.e. the probabilities and latent fields for location  $s$  are given by  $\{z_{s,k}\}_{k=1}^D$  and  $\{\eta_{s,k}\}_{k=1}^d$ , respectively.

In this paper the link function is constructed by applying the additive log-ratio (alr) transform

$$\eta_{s,k} = \log z_{s,k} - \log z_{s,D}, \quad k = 1, \dots, D - 1 \quad (3)$$

for each location  $s$ .

## 2.2 Latent Field

Given a total of  $N \geq N_o$  locations at which we want to provide composition predictions the latent field,  $\boldsymbol{\eta}_{all}$ , is multivariate with  $d = D - 1$  elements at each location ( $N \geq N_o$  since we are providing predictions at the observed and some additional locations). To simplify notation the latent field is represented as a  $Nd \times 1$  vector  $\boldsymbol{\eta}_{all} = (\eta_{all,1}^\top, \dots, \eta_{all,d}^\top)^\top$ , where each  $\eta_{all,k}$  is spatial field with  $N$  locations.

The latent field and its connection to the observed locations is given as:

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{A}\boldsymbol{\eta}_{all} \\ \boldsymbol{\eta}_{all} &= \mathbf{B}\boldsymbol{\beta} + \mathbf{X}. \end{aligned} \quad (4)$$

where  $\mathbf{A} = \mathbb{I}_{d \times d} \otimes \mathbf{A}$  which extracts observed elements from  $\boldsymbol{\eta}_{all}$  with  $\mathbf{A}$  being a  $N_o \times N$  sparse observation matrix,  $\mathbf{B} = \mathbb{I}_{d \times d} \otimes \mathbf{B}$  with  $\mathbf{B}$  being a  $N \times p$  matrix of covariates,  $\boldsymbol{\beta}$  is a  $dp \times 1$  matrix of regression coefficients, and  $\mathbf{X} = (X_1^\top, \dots, X_d^\top)^\top$  is a spatially correlated multivariate field. With this structure, the spatial dependence  $\mathbf{X}$ , can be modelled as a GMRF with a separable covariance structure, i.e.  $\boldsymbol{\rho} \otimes \mathbf{Q}^{-1}$ , which captures the dependency among and within the fields;

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\rho} \otimes \mathbf{Q}^{-1}(\kappa)). \quad (5)$$

Here  $\boldsymbol{\rho}$  is a  $d \times d$  matrix of covariances among the  $d$  multivariate fields ( $X_k$ ,  $k = 1, \dots, d$ ), and  $\mathbf{Q}(\kappa)$  is a  $N \times N$  precision matrix of a GMRF with spatial scale parameter  $\kappa$ .  $\mathbf{Q}$  is chosen as the precision matrix which approximates a stationary Matérn field with smoothness  $\nu = 1$  (see Lindgren et al., 2011, for details); this precision is also a solution to a stationary stochastic partial differential equation field (SPDE) with  $\alpha = 2$ .

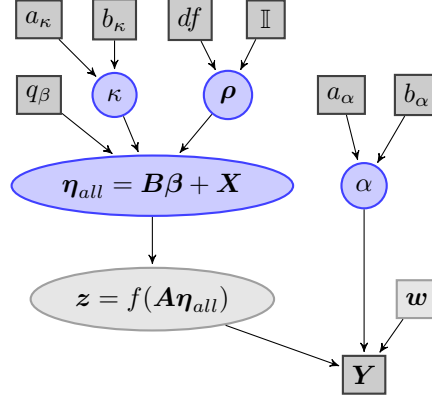


Figure 1: Directed acyclic graph describing the conditional dependencies on the hierarchical model.

### 2.3 Hierarchical Model and Priors

The full hierarchical model (Fig. 1) based on Dirichlet observations (1) of a transformed latent GMRF (5) becomes

$$\begin{aligned}
 \mathbf{y}_s | \alpha, \boldsymbol{\eta} &\sim \text{Dir}(\alpha \mathbf{w}_s f_s(\boldsymbol{\eta})), & s = 1, \dots, N_o \\
 \boldsymbol{\eta}_{all} &= \mathbf{B}\boldsymbol{\beta} + \mathbf{X}, & \boldsymbol{\eta} = \mathbf{A}\boldsymbol{\eta}_{all}, \\
 \mathbf{X} | \kappa, \boldsymbol{\rho} &\sim \mathcal{N}(\mathbf{0}, (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa))^{-1}), & (6) \\
 \boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, \mathbb{I} q_\beta^{-1}), & \alpha \sim \Gamma(a_\alpha, b_\alpha) \\
 \kappa &\sim \Gamma(a_\kappa, b_\kappa), & \boldsymbol{\rho} \sim IW(\mathbb{I}, df)
 \end{aligned}$$

To make  $\mathbf{X}$  and  $\boldsymbol{\beta}$  jointly normal, we use a vague Gaussian prior for  $\boldsymbol{\beta}$  with precision  $q_\beta = 10^{-3}$ . The Dirichlet scale parameter,  $\alpha$ , and spatial scale parameter,  $\kappa$ , are given gamma priors, and for  $\boldsymbol{\rho}$  we choose a conjugate prior for covariance matrices, the inverse Wishart ( $IW$ ). Values of all hyper parameters are given below:

$$\begin{aligned}
 a_\alpha &= 1.5, & b_\alpha &= 0.1, & \mathbb{I} &= d \times d \text{ identity matrix,} \\
 a_\kappa &= 1.5, & b_\kappa &= 0.1, & df &= 10.
 \end{aligned}$$

Having detailed the model, parameter estimation and reconstruction of  $\boldsymbol{\eta}_{all}$ , using MCMC, are described in the following section.

## 3 Estimation Using MCMC

A block-updated MCMC algorithm is used to estimate the latent field  $\boldsymbol{\eta}_{all}$  and the unknown parameters  $\alpha, \kappa, \boldsymbol{\rho}$ . For GMRF, joint updating of parameters in as large blocks as possible has been shown to improve mixing and convergence (Knorr-Held and Rue, 2002). Therefore, the algorithm in this paper updates the unknowns by alternating between two blocks: the first block updates the latent fields and the Dirichlet scale parameter using MALA (Girolami and Calderhead, 2011), the second block updates the parameters of the GMRF,  $\kappa$  and  $\boldsymbol{\rho}$ , using a combination of random walk proposals and the conjugate posterior for  $\boldsymbol{\rho}$ .

### 3.1 Updating $\boldsymbol{\eta}$ and $\alpha$

To update  $\boldsymbol{\eta}$  and  $\alpha$  we use a Metropolis-Hastings (MH) step to draw samples from the conditional distribution

$$\begin{aligned} \mathbb{P}(\boldsymbol{\eta}, \alpha | \boldsymbol{\kappa}, \boldsymbol{\rho}, \mathbf{Y}) &\propto \left( \prod_{s=1}^{N_o} \mathbb{P}(\mathbf{y}_s | f(\boldsymbol{\eta}), \alpha) \right) \cdot \mathbb{P}(\mathbf{X} | \boldsymbol{\kappa}, \boldsymbol{\rho}) \cdot \mathbb{P}(\boldsymbol{\beta}) \cdot \mathbb{P}(\alpha) \\ &\propto \prod_{s=1}^{N_o} \left( \frac{\Gamma(\alpha w_s)}{\prod_{k=1}^D \Gamma(\alpha w_s z_{s,k})} \prod_{k=1}^D y_{s,k}^{\alpha w_s z_{s,k} - 1} \right) \\ &\quad \cdot \exp \left( -\frac{1}{2} \mathbf{X}^\top (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\boldsymbol{\kappa})) \mathbf{X} \right) \\ &\quad \cdot \exp \left( -\frac{q\boldsymbol{\beta}}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} \right) \\ &\quad \cdot \alpha^{a_\alpha - 1} e^{-\alpha \cdot b_\alpha}. \end{aligned} \quad (7)$$

The MH-step uses a MALA proposal:

$$\boldsymbol{\eta}^*, \alpha^* | \boldsymbol{\eta}, \alpha \quad \sim \quad \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\eta} \\ \alpha \end{bmatrix} + \frac{\epsilon^2}{2} \mathbf{FI}^{-1} \nabla l, \epsilon \mathbf{FI}^{-1} \right), \quad (8)$$

where  $\epsilon$  is the step size of MALA,  $\nabla l$  is a vector of derivatives of  $\log \mathbb{P}(\boldsymbol{\eta}, \alpha | \boldsymbol{\kappa}, \boldsymbol{\rho}, \mathbf{Y})$  w.r.t.  $\boldsymbol{\eta}$  and  $\alpha$  (for computational details see Appendix A.2) and  $\mathbf{FI}$  is the expected Fisher information matrix, (see Appendix A.3). At each iteration,  $\mathbf{FI}^{-1} \cdot \nabla l$  gives a sampling direction from the current state which is similar to a Newton-Raphson step (Givens and Hoeting, 2012, Ch. 2). Further, the proposal variance,  $\mathbf{FI}^{-1}$ , accounts for the dependency among the parameters. Due to the GMRF structure of the latent fields,  $\mathbf{FI}$  will be a sparse matrix reducing the computations to sampling from a GMRF, for which efficient algorithms exist (Rue and Held, 2004).

In order to get reasonable acceptance rate, an adaptive MCMC method (Andrieu and Thoms, 2008) is used for  $\epsilon$  with the following updating rule;

$$\epsilon_{i+1} = \epsilon_i + \gamma_{i+1} (\widehat{acc}_{\boldsymbol{\eta}, \alpha}(\epsilon_i) - 0.57) \quad (9)$$

where  $\epsilon_i$  is the step size of the  $i^{\text{th}}$  MCMC iteration,  $\gamma_i = i^{-1/2}$ ,  $\widehat{acc}$  is the acceptance probability of the  $i$ th step, and 0.57 is the target acceptance rate as suggested in Roberts and Rosenthal (1998).

### 3.2 Updating $\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$

The second block is updated using a combination of the conjugate posterior for  $[\boldsymbol{\rho} | \mathbf{X}, \boldsymbol{\kappa}]$  and a M-H random walk (in log scale) for  $[\boldsymbol{\kappa} | \mathbf{X}]$ . The joint posterior of  $[\boldsymbol{\kappa}, \boldsymbol{\rho} | \mathbf{X}]$  can be written as

$$\mathbb{P}(\boldsymbol{\kappa}, \boldsymbol{\rho} | \mathbf{X}) = \mathbb{P}(\boldsymbol{\rho} | \mathbf{X}, \boldsymbol{\kappa}) \cdot \mathbb{P}(\boldsymbol{\kappa} | \mathbf{X}). \quad (10)$$

Due to the conjugate prior for  $\boldsymbol{\rho}$  the conditional posterior for  $\boldsymbol{\rho}$  is inverse Wishart;

$$[\boldsymbol{\rho} | \boldsymbol{\kappa}, \mathbf{X}] \propto IW(\mathbb{I} + \mathbf{x}^\top \mathbf{Q}(\boldsymbol{\kappa}) \mathbf{x}, N + df) \quad (11)$$

with  $\mathbf{x}$  being a  $N \times d$  matrix given by  $\mathbf{x} = [X_1, \dots, X_d]$ . The conjugacy makes it possible to marginalize over  $\boldsymbol{\rho}$  (see Appendix B) giving

$$\mathbb{P}(\kappa|\mathbf{X}) \propto \int \mathbb{P}(\boldsymbol{\rho}|\kappa, \mathbf{X}) \mathbb{P}(\kappa) d\boldsymbol{\rho} \propto \frac{|\mathbf{Q}(\kappa)|^{\frac{D}{2}}}{|\mathbb{I} + \mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x}|^{\frac{N+df}{2}}} \mathbb{P}(\kappa). \quad (12)$$

Samples from  $[\kappa, \boldsymbol{\rho}|\mathbf{X}]$  are now obtained by first sampling from the posterior (12) using a MH-step random-walk proposal in log scale,

$$\log \kappa^* = \log \kappa + \epsilon_\kappa, \quad \epsilon_\kappa \sim \mathcal{N}(0, \sigma_\kappa^2).$$

If the proposed  $\kappa^*$  is accepted, a new proposal,  $\boldsymbol{\rho}^*$ , is sampled from (11). Using these updates can be seen as a joint MH-step for  $\boldsymbol{\rho}$  and  $\kappa$  with proposal density  $q(\kappa^*, \boldsymbol{\rho}^*|\kappa) \sim \mathbb{P}(\boldsymbol{\rho}^*|\mathbf{X}, \kappa^*) q(\kappa^*|\kappa)$  and acceptance ratio:

$$\text{acc}_{\kappa, \boldsymbol{\rho}} = \min \left( 1, \frac{\mathbb{P}(\kappa^*, \boldsymbol{\rho}^*|\mathbf{X})}{\mathbb{P}(\kappa, \boldsymbol{\rho}|\mathbf{X})} \cdot \frac{q(\kappa, \boldsymbol{\rho}|\kappa^*)}{q(\kappa^*, \boldsymbol{\rho}^*|\kappa)} \right) = \min \left( 1, \frac{\mathbb{P}(\kappa^*|\mathbf{X})}{\mathbb{P}(\kappa|\mathbf{X})} \cdot \frac{\kappa^*}{\kappa} \right). \quad (13)$$

The proposal variance,  $\sigma_\kappa^2$ , is determined using an adaptive scheme similar to (9), with target acceptance rate of 0.3.

## 4 Uncertainty

To obtain uncertainties in the composition estimates at each location, we use the MCMC samples of  $\boldsymbol{\eta}$  at each location. Given the model structure with a Gaussian prior for  $\boldsymbol{\eta}$  we base the joint confidence regions (CRs) for the composition estimates on the elliptical CRs obtained for multivariate Gaussian distributions. Using the sample mean,  $\boldsymbol{\mu}$ , and the sample covariance,  $\boldsymbol{\Sigma}$ , in the MCMC samples, we construct the CR for each location as the ellipse

$$(\boldsymbol{\eta} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}) = C_\alpha. \quad (14)$$

The quantile  $C_\alpha$  is taken as the  $\alpha$ -quantile of the above squared Mahalanobis distance computed for all the MCMC samples (for a multivariate Gaussian  $C_\alpha = \chi_\alpha^2(d)$ ). Thereafter, the confidence ellipse is transformed from  $R^d$  to  $(0, 1)^D$  using (2). For illustration purposes, we choose  $D = 3$ . The new ternary region is considered as 95% CR for the transformed  $\boldsymbol{\eta}$ , i.e. the composition estimates, see Fig. 2.

To illustrate the changes in compositions, we choose the maximum and minimum along each dimensions of the ternary plot, i.e. in each component. This way, we get a joint lower bound (minimum) and upper bound (maximum) for each composition together with the corresponding changes in the other compositions “most likely” to occur at the bounds (Fig. 2).

In addition, we compute prediction regions (PRs) for the compositions. To obtain PRs, we simulate new Dirichlet observations for each MCMC sample of  $\boldsymbol{\eta}$  and  $\alpha$ . These D-composition Dirichlet-simulations are then transformed to  $\mathbb{R}^d$  using the link function. The procedure above is then used to obtain prediction ellipses and ternary PRs.

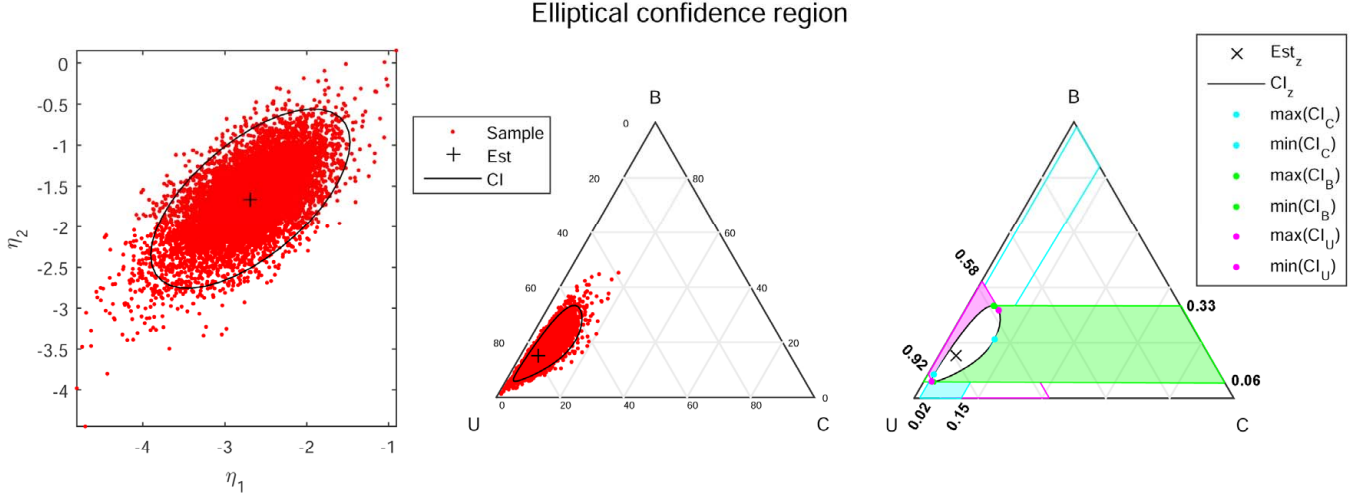


Figure 2: The left plot shows the 95% elliptical confidence region for the  $\eta$  samples. The middle ternary diagram shows the transformed samples and ellipse. The right hand ternary diagram shows the joint maximum and minimum in each composition, C, B, and U; together with the confidence interval for each compositions.

## 5 Application

The model was tested on simulated data (Section 5.1) before being applied to the PbLCC data (Section 5.2). For both cases two version of the model in (4) were used: 1) a Full spatial model with  $\eta_{all} = \mathbf{B}\beta + \mathbf{X}$  (includes all parameters,  $\alpha, \beta, \rho, \kappa$  and the  $\mathbf{X}$  fields), and 2) a regression model (RM) with no spatial structure where  $\eta_{all} = \mathbf{B}\beta$ . The intrinsic GMRF (IGMRF) model ( $\kappa = 0$ ), also considered in Pirzamanbein et al. (2014), performed similar to or slightly worse than the Full model with Dirichlet observations, hence those results have been excluded for brevity.

### 5.1 Simulated Data

To investigate the model performance we applied the two models explained in section 5 to simulated data. The data consisted of 400 observed locations randomly distributed over a  $40 \times 30$  lattice. To emulate the three land cover-classes in the LCC data we took  $D = 3$ , giving two latent fields in  $\eta$ . The covariate matrix,  $\mathbf{B}$ , consisted of an intercept and a random vector of i.i.d.  $\mathcal{N}(0, 1)$  numbers; the weight vector,  $\mathbf{w}$ , was made up of i.i.d.  $U(0.5, 2)$  numbers; the latent fields and observations were simulated using the parameters (True) in Table 1.

Given the simulated observations we ran 100 000 MCMC iterations with a burn-in of 10 000 to estimate the parameters of each model. Autocorrelation plots showed good mixing of the parameters. To investigate the role of  $\mathbf{w}$  we estimated the two models using both the simulated  $\mathbf{w}$  and a constant  $\mathbf{w} \equiv 1$ . Using the true  $\mathbf{w}$  values lead to better estimates for both models; for brevity we only report the results with  $\mathbf{w} \equiv 1$  for the Full model.



Table 1 presents the parameter estimates of the MCMC chains for the two models. Note that,  $\rho_{hl}$  indicates the variance of each field when  $h = l$ , and the covariance between the fields when  $h \neq l$ . Since  $\boldsymbol{\rho}$  is symmetric, only the upper triangular part is reported. The first index of  $\beta$  represents different fields and the second index represents the covariates, e.g.  $\beta_{10}$  indicates the intercept coefficient of field 1.

Table 1: True values, parameter estimates (Est) and 95% quantiles (CI) for the simulated data for Full, Full<sub>w=1</sub> and RM.

Parameter	True	Full		Full <sub>w=1</sub>		RM	
		Est	(CI)	Est	(CI)	Est	(CI)
$\alpha$	3	2.93	( 2.55 , 3.36 )	3.92	( 3.35 , 4.58 )	0.78	( 0.72 , 0.84 )
$\kappa$	0.3	0.19	( 0.12 , 0.27 )	0.23	( 0.15 , 0.32 )		
$\rho_{11}$	0.25	0.72	( 0.41 , 1.23 )	1.04	( 0.57 , 1.75 )		
$\rho_{12}$	-0.1	-0.12	( -0.30 , 0.09 )	-0.07	( -0.35 , 0.25 )		
$\rho_{22}$	0.16	0.42	( 0.23 , 0.75 )	0.72	( 0.37 , 1.33 )		
$\beta_{10}$	0.5	0.58	( -0.90 , 2.21 )	0.56	( -0.68 , 1.73 )	0.26	( 0.16 , 0.37 )
$\beta_{11}$	1	1.01	( 0.89 , 1.14 )	1.06	( 0.93 , 1.19 )	0.83	( 0.73 , 0.93 )
$\beta_{20}$	0.5	0.59	( -0.64 , 1.78 )	0.66	( -0.34 , 1.68 )	0.27	( 0.17 , 0.37 )
$\beta_{21}$	1	1.00	( 0.88 , 1.12 )	1.04	( 0.91 , 1.16 )	0.74	( 0.64 , 0.85 )

The estimated parameters in Table 1 show that the Full model is able to recover the true parameters used for simulation; the lack of known weights in Full<sub>w=1</sub> or the lack of spatial structure in the RM leads, as expected, to worse estimates.

The simulated data and the results from the Full model are illustrated in Fig. 3. The model's predictive performance was investigated using the average compositional distances (ACD) between the simulated  $\mathbf{z}$  fields and the MCMC reconstructions. The compositional distances (Aitchison et al., 2000) were computed for each location, using

$$\text{ACD}(\mathbf{u}, \mathbf{v}) = [(\mathbf{u} - \mathbf{v})^T \mathbf{J}^{-1}(\mathbf{u} - \mathbf{v})]^{1/2} \quad (15)$$

where  $\mathbf{u}$  and  $\mathbf{v}$  are alr transforms of the compositions and  $\mathbf{J}$  is a  $d \times d$ -matrix with elements  $J_{h,l} = 2$  if  $h = l$ , and  $J_{h,l} = 1$  if  $h \neq l$ , and averaged over all locations. We also computed the root mean square error (RMSE) between simulated  $\boldsymbol{\eta}$  fields and model estimates, see Table 2. These comparisons show that the Full model, with known weights, is highly capable of recovering the unobserved latent fields. The RM of course only captures the part of the latent field attributable to the covariates, and the larger difference between the latent field and the observations gives a larger spread in the Dirichlet observations (i.e. smaller alpha estimate).

Table 2: The first row shows the average compositional distance (ACD) between the simulated  $\mathbf{z}$  fields (compositions) and the MCMC reconstructions; the second row shows the root mean square error (RMSE) between the simulated  $\boldsymbol{\eta}$  fields (latent GMRFs) and the MCMC reconstructions.

	Full	Full <sub><math>w \equiv 1</math></sub>	RM
ACD	<b>0.4951</b>	0.5233	1.2627
RMSE	<b>0.5395</b>	0.5723	1.1802

Simulated data and Full model reconstruction

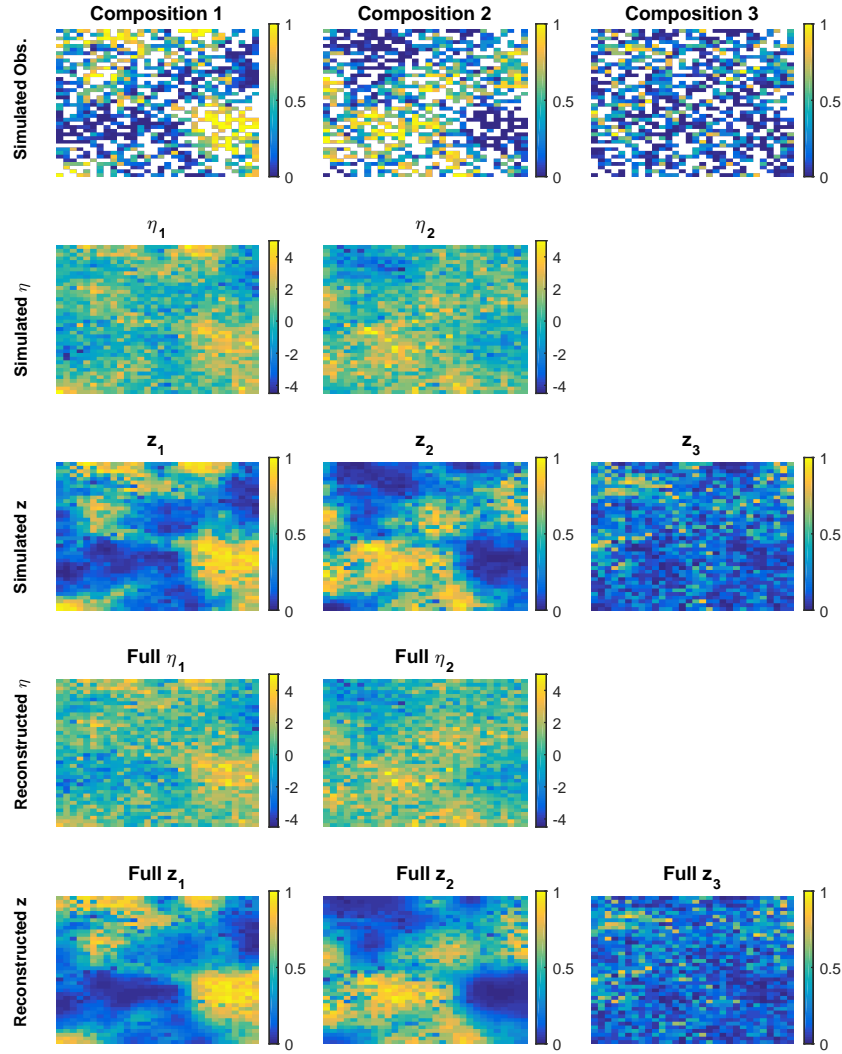


Figure 3: Rows 1-3 show the simulated observations,  $\boldsymbol{\eta}$  fields, and  $\mathbf{z}$  fields, respectively. Row 4 and 5 show the reconstructed fields using the Full model.

## 5.2 Land-cover Compositions

To reconstruct and interpolate the PbLCC data we considered the two models used for the synthetic data, but with different covariates. The remainder of this section consists of: a description of the data (Sec. 5.2.1), parameters estimation and spatial reconstruction results for the models (Sec. 5.2.2), and validation of model performance (Sec. 5.2.3).

### 5.2.1 Data

The data used for the reconstruction of past LCC over Europe consists of pollen-based REVEALS estimates (here called pollen-based LCC data — PbLCC) of the three LCTs: Coniferous forest (C), Broadleaved forest (B) and Unforested land (U). The PbLCC data was obtained using the REVEALS model (Sugita, 2007b); a mechanistic model that takes into account the size of sedimentary basins and inter-taxonomic differences in pollen productivity and dispersal to estimate vegetation cover from pollen records at regional scales. The REVEALS model provides PbLCC estimates for each  $1^\circ \times 1^\circ$  grid cell containing lakes or bogs for which pollen records have been recovered (Hellman et al., 2008). The  $1^\circ \times 1^\circ$  grid is an appropriate scale for climate models, which currently work at this or higher resolutions (Trondman et al., 2015). Although the REVEALS model provides uncertainties, the uncertainties are given for each individual vegetation species in the LCTs. Since no measure of correlation between species is given, the species level uncertainties cannot be aggregated to relative uncertainties for each location. Hence, a constant weight parameter for all locations ( $w \equiv 1$ ) is used in the LCC reconstruction.

The PbLCC data are available for five time periods centred around 1900, 1725 and 1425 CE, 1000 and 4000 BCE, with 175, 181, 193, 204 and 196 observed grid cells, respectively (Trondman et al., 2015). These time periods are commonly used in both climate modelling and palaeoecological studies since they represent major climatic and historical events; Recent Past, Little Ice Age, Black Death, Late Bronze Age, and Early Neolithic.

To capture large scale structures in the LCC, covariates consisting of potential natural vegetation cover adjusted for human land use, and elevation were used. The choice of covariates was based on the best model found in Pirzamanbein et al. (2014), and detailed descriptions of the covariates can be found in that paper. Here we only provide a brief summary.

DVM based estimates of climate-induced potential natural vegetation, for the study area and specified time periods, were obtained using the LPJ-GUESS model (Smith et al., 2001). To account for human land use, the potential natural vegetation was adjusted for anthropogenic deforestation using the KK10 scenarios of Kaplan et al. (2009). The KK10 scenarios provide assessments of human induced deforestation based on estimates of past human population densities, land area required for food production to sustain that population, and a model of land suitability for food production. Combining the potential natural vegetation cover from LPJ-GUESS and the KK10 scenarios of deforestation resulted in a land cover covariate, denoted  $\text{LPJ-GUESS}_{\text{KK}}$ .

The elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) (Becker et al., 2009)<sup>1</sup> upscaled and truncated to  $\geq 0$  to avoid a few

<sup>1</sup>downloaded from [ftp://topex.ucsd.edu/pub/srtm30\\_plus/](ftp://topex.ucsd.edu/pub/srtm30_plus/) on 20110903

grid cells along the Norwegian coast with average elevation down to 1000 metres below sea level.

Since the potential land cover,  $\text{LPJ-GUESS}_{\text{KK}}$ , is compositional it was transformed using (2), and the covariate matrix,  $\mathbf{B}$  consisted of the following columns:  $B_0$  – intercept;  $B_1, B_2$  – alr-transformed  $\text{LPJ-GUESS}_{\text{KK}_{1,2}}$ ; and  $B_3$  – elevation.

To evaluate our results we used present-time European forest maps compiled by the European Forest Institute (EFI). These maps are based on a combination of satellite data (NOAA-AVHRR) and national forest-inventory statistics from 1990–2005 (Päivinen et al., 2001; Schuck et al., 2002)<sup>2</sup>. The EFI forest maps (EFI-FM; with proportions of coniferous- and broadleaved-forest cover) were upscaled from 1 km  $\times$  1 km to 1°  $\times$  1° resolution. The proportions of unforested area were calculated by subtracting the total sum of forested cover from 1.

### 5.2.2 Results

To estimate the parameters for each model, we ran 100 000 MCMC iterations with a burn-in sample size of 10 000. Diagnostics for the chains indicate a fast convergence for  $\alpha, \rho$  and  $\beta$ ; autocorrelation plots show good mixing of all parameters after burn-in.

Parameter estimates for the 1900 CE time period are given in Table 3; the parameter estimations for the other time periods can be found in Appendix C. Note that the  $\alpha$  estimate for the RM is lower than for the Full model, indicating higher observational variation in RM.

Table 3: Parameter estimates (Est) and 95% confidence intervals (CI) for the two models fitted to the PbLCC data from the 1900 CE time period.

1900 CE				
Parameter	Full		RM	
	Est	(CI)	Est	(CI)
$\alpha$	10.86	( 8.10 , 15.41 )	6.36	( 5.58 , 7.18 )
$\kappa$	0.28	( 0.14 , 0.45 )	-	-
$\rho_{11}$	0.78	( 0.12 , 2.61 )	-	-
$\rho_{12}$	0.57	( 0.05 , 1.96 )	-	-
$\rho_{22}$	0.60	( 0.10 , 1.97 )	-	-
$\beta_{10}$	-0.68	( -1.64 , 0.15 )	-0.13	( -0.25 , -0.02 )
$\beta_{11}$	0.16	( 0.08 , 0.24 )	0.24	( 0.22 , 0.27 )
$\beta_{12}$	0.02	( -0.09 , 0.14 )	-0.03	( -0.09 , 0.02 )
$\beta_{13}$	0.05	( -0.15 , 0.26 )	-0.10	( -0.19 , -0.01 )
$\beta_{20}$	-0.94	( -1.83 , -0.22 )	-0.38	( -0.51 , -0.26 )
$\beta_{21}$	0.04	( -0.05 , 0.12 )	0.13	( 0.11 , 0.16 )
$\beta_{22}$	0.01	( -0.09 , 0.11 )	-0.05	( -0.10 , 0.00 )
$\beta_{23}$	-0.04	( -0.24 , 0.16 )	-0.24	( -0.34 , -0.14 )

Reconstructions of the LCC for the two models and the 1900 CE time period are shown in Fig. 4. Results for the other time periods are available in Appendix D. Figure 4 shows that the LCC reconstructions from the two models captured

<sup>2</sup>downloaded from the EFI webpage [http://www.efi.int/portal/virtual\\_library/information\\_services/mapping\\_services/forest\\_map\\_of\\_europe](http://www.efi.int/portal/virtual_library/information_services/mapping_services/forest_map_of_europe)

the structure in the PbLCC data. However, the results from RM is smoother than from the Full model. The Full model better captures the high abundance of unforested land in Poland, Denmark and south east Norway.

The uncertainties in the LCC reconstructions were computed using the method described in Sec. 4. Results for the 1900 CE time period are presented in Fig. 5 and 6, with results for the remaining time periods given in Appendix E. The confidence and prediction regions represent the uncertainty in the latent field reconstruction,  $\mathbf{z}_s$  and the potential uncertainty in new PbLCC data,  $\mathbf{y}_s$ , for a given grid cell, respectively. In general the Full model has larger CRs but smaller PRs than RM (Fig. 6). This is due to the spatial component in the Full model being able to better capture spatial variation resulting in a lower uncertainty (larger  $\alpha$ ) in the Dirichlet observations as compared to RM. The maps of CRs (Fig. 5) illustrate rather large uncertainties in the predicted LCC in general, and especially for Southeast Europe, a region with very few observations.

### 5.2.3 Validation

To evaluate the performance of the models, we compared the LCC reconstructions for 1900 CE to the EFI-FM by computing the ACD using (15). Although a temporal misalignment exists between the PbLCC data (PbLCC data are from 1850 to recent years) and the EFI-FM (inventory and satellite data are from 1990-2005); EFI-FM provides the best complete and consistent land cover map of Europe for present times, making it a reasonable choice for the comparison. Figure 4 shows the maps of PbLCC data and EFI-FM. The main differences between the EFI-FM data and the PbLCC data for the 1900 CE time period are: 1) a lower abundance of broadleaved forest around most of Europe, 2) a higher abundance of coniferous forest in Sweden and Finland, and 3) a higher abundance of unforested land in North Norway in the EFI-FM data than in the PbLCC data. Compositional distances between LCC reconstructions and EFI-FM were computed using (15) and averaged over all grid cells. The resulting ACD are 1.4757 and 1.5025 for Full model and RM, respectively. This indicates that the Full model provides a reconstruction closer to EFI-FM than RM.

These results can also be compared to a model with Gaussian observations of transformed latent fields, (Pirzamanbein et al., 2014). The resulting ACD of the Gaussian observation models compare to the EFI-FM are  $ACD_{IGMRF} = 1.6007$  and  $ACD_{RM} = 1.6140$ . The differences between what Pirzamanbein et al. (2014) reported ( $ACD_{IGMRF} = 1.5201$  and  $ACD_{RM} = 1.5177$ ) and our results using their models are due to an increase in available data leading to more grid cells in our reconstructions. These results indicate smaller distances between the LCC reconstructions and EFI-FM for the models with Dirichlet observations proposed in this paper compared to similar models with Gaussian observations.

Since no ground truth exists for the other time periods, we applied a 6-fold cross-validation scheme for the models for each of the five time periods (Friedman et al., 2001, Ch. 7.10). The cross-validation was run for 10 different, randomly selected 6 folds to assess the variability due to different cross validation groupings. Average compositional errors and standard deviations are shown in Table 4. The Full model gives the best predictions for all the five time periods.

Table 4: Average compositional error (and standard deviation) from 10 different 6-fold cross-validations for each of the models, and time periods.

Time	Full		RM	
	$CV_{\text{error}}$ (sd)		$CV_{\text{error}}$ (sd)	
1900 CE	<b>1.0169</b>	(0.0122)	1.1439	(0.0061)
1700 CE	<b>1.1448</b>	(0.0084)	1.2891	(0.0054)
1400 CE	<b>1.2009</b>	(0.0071)	1.4061	(0.0042)
1000 BCE	<b>1.3260</b>	(0.0083)	1.5287	(0.0062)
4000 BCE	<b>1.2131</b>	(0.0109)	1.3396	(0.0045)

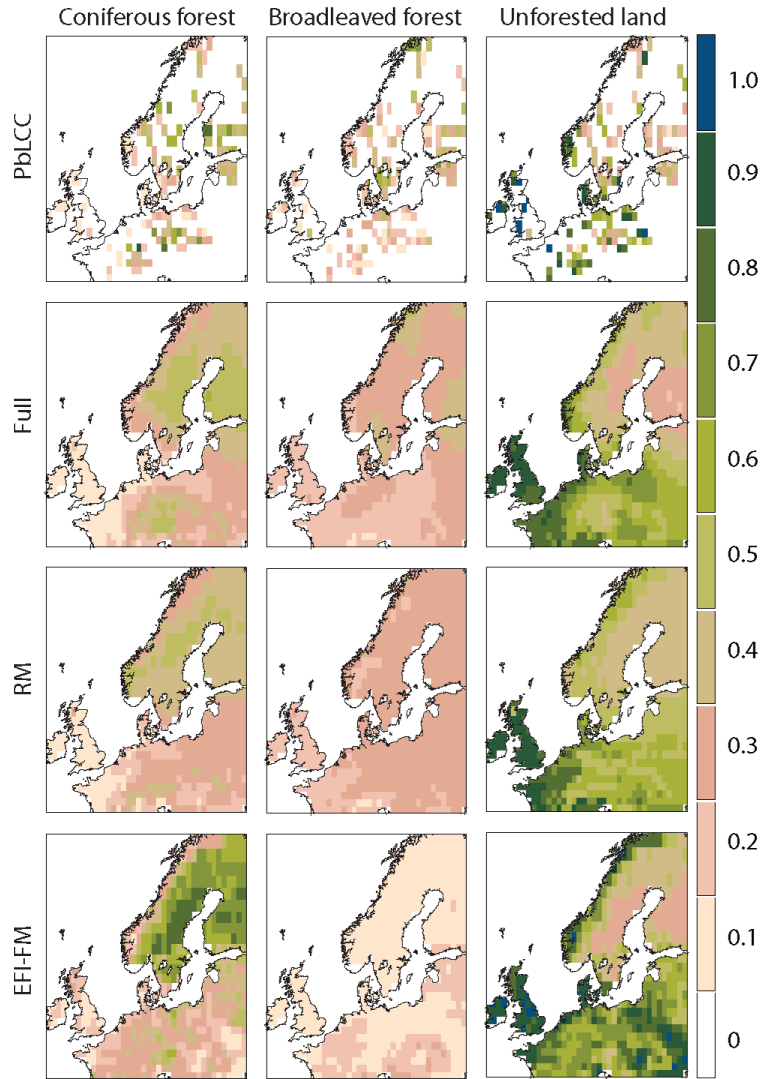


Figure 4: Results for the 1900 CE time period: the top row shows the PbLCC data from REVEALS, the bottom row shows the EFI-FM and the remaining rows show the reconstructions for the Full model and RM.

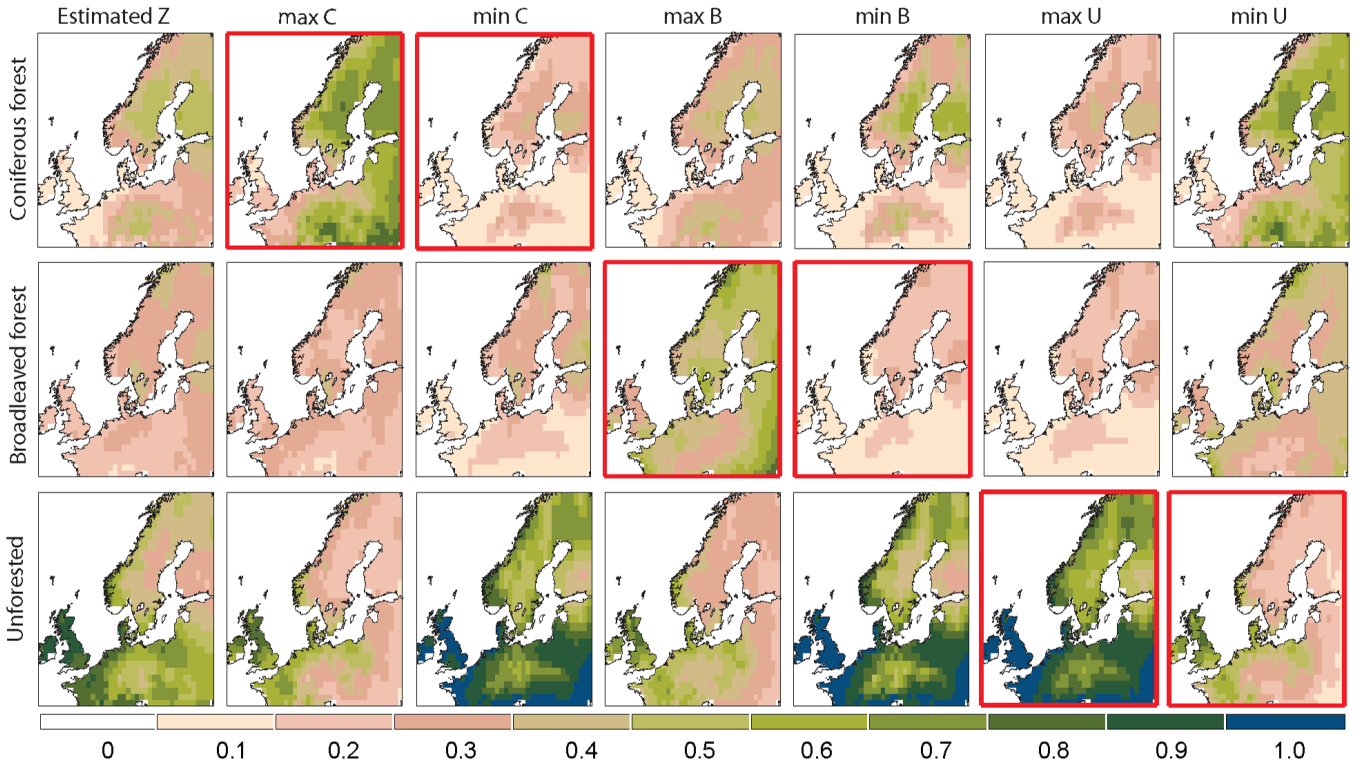


Figure 5: The first column shows the reconstructed LCC for the 1900 CE time period. Rows 2 and 3, column 1 (with tick axes), show the maximum and minimum of 95% elliptical CRs for Coniferous; columns 2 and 3 give the corresponding Broadleaved and Unforested compositions. Rows 4 and 5 (column 2 with tick axes) gives the bounds for the Broadleaved composition while rows 6 and 7 show the bounds for Unforested land (column 3 with tick axes). For each of the rows the corresponding changes in the remaining two compositions are also given.

# Confidence and Prediction region 1900 CE

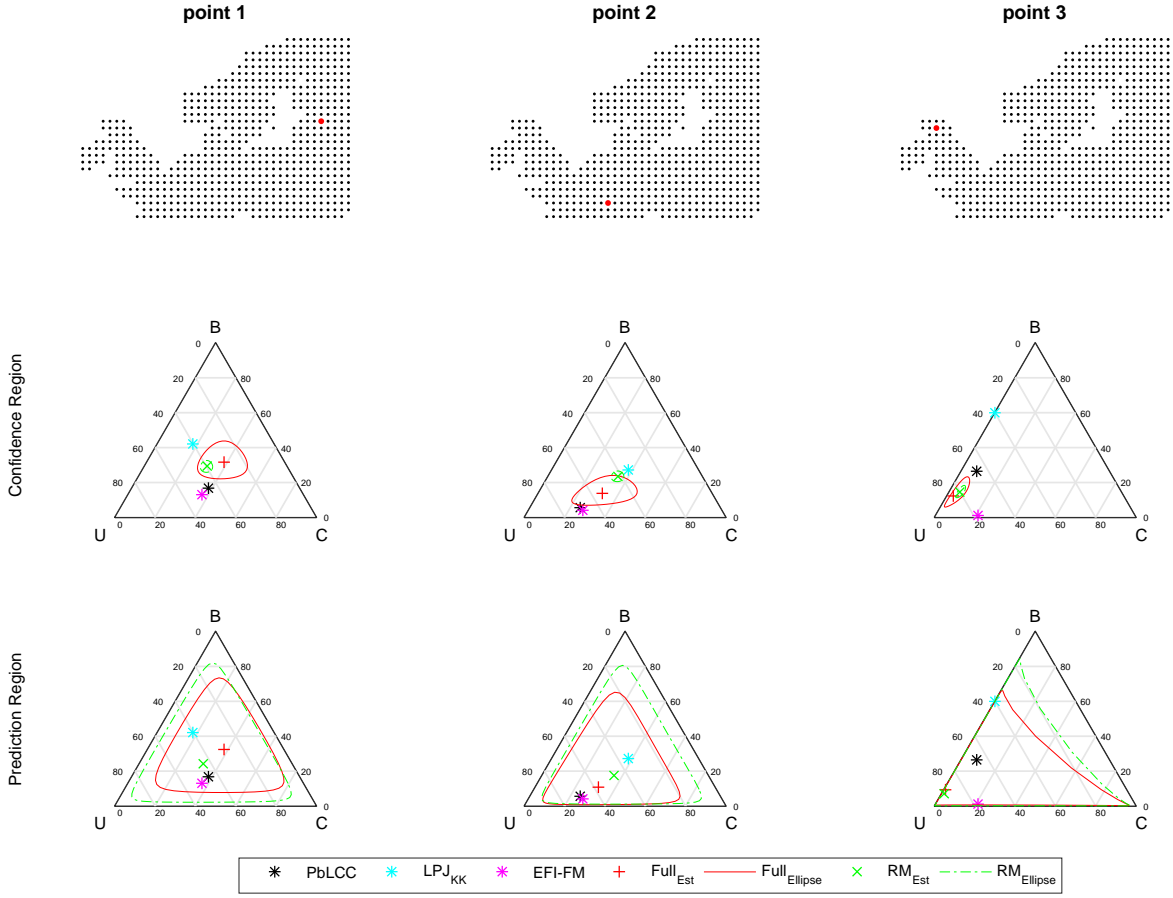


Figure 6: The first row shows the locations of the three selected grid cells. The second row shows the ternary CRs and the LCC reconstructions for the two models together with the PbLCC data from REVEALS, the LPJ-GUESS<sub>KK</sub> land cover covariate and the EFI-FM for each location. The third row shows the ternary PRs.



## 6 Conclusion

In this paper we have introduced a model for spatial interpolation of compositional data that relies on Dirichlet observations of an underlying multivariate GMRF. In theory the formulation allows for a wide class of link-functions between the GMRF and the compositional probabilities in the Dirichlet observations; we used the  $\text{alr}$ -transformation throughout the paper. Since the sparse structure in the precision matrix of the GMRF carries over to the expected Fisher information used in MALA, the model formulation with a latent GMRF allows for fast MCMC-based estimation of parameters and latent field. As a result our MCMC produced 200 samples per second using MATLAB® on a standard desktop (Intel® Core™ i7 – 2600 CPU (2011) with 8 GB memory) for a latent field with 2160 nodes (bivariate field on a 27-by-40 grid).

The model and MCMC-based parameter estimation scheme were evaluated using simulated data, for which we are able to recover the true parameters. The model is capable of incorporating known relative differences in the uncertainty of the observed compositional data at different locations. Results from the simulation study show improved parameter estimation and reconstruction of the latent field when these differences are included in the model. Due to lack of relevant uncertainty information in the PbLCC data we are currently unable to use this model feature for the real data, but we are certain it will prove useful in the future.

To evaluate prediction uncertainties we also proposed a method for construction of joint confidence and prediction regions of the predicted compositions at each location. The idea behind the method is to use the MCMC samples to first construct elliptical CRs for the transformed latent fields; these are then transformed from  $R^d$  to  $(0, 1)^D$  using the inverse link-function, giving CRs in compositional space. Having joint CRs for the compositions allowed us to evaluate the behaviour of all components as each individual component attains their lower and upper bounds in the CRs.

The statistical model was used to reconstruct past LCC over Europe for five time periods using PbLCC data (Trondman et al., 2015) obtained from the REVEALS model (Sugita, 2007b). The LCC reconstructions for the most recent time period were evaluated against modern forest maps, and reconstructions for all time-periods were evaluated using cross-validation. The evaluations showed that a model containing both explanatory covariates and spatial dependence structure outperformed a model with only covariates. Evaluations using the modern forest maps showed that a model with Dirichlet observations outperformed previously developed models using Gaussian observations of transformed fields (Pirzamanbein et al., 2014).

The reconstructed maps of LCCs can be used both in studies of climate models and to analyse changes in LCCs during the past millennia. For example, Fig. 7 uses the compositional distances (15) to illustrate the changes in LCC between the five time periods considered in this study. This simple analysis shows that the largest changes in land cover between 4000 BCE and 1900 CE have occurred in Switzerland and Central France; along the North Sea coast in the UK, the Low Countries, Denmark, and southern Norway; and along the south Baltic coast in northern Germany and Poland.

The reconstructions of past LCC obtained here are encouraging, as they clearly show the ability to recover continuous maps of past land cover from

PbLCC data. The reconstructions from the Full model appear to conserve the information and trends from the pollen-based REVEALS estimates of past land cover (as discussed in Trondman et al., 2015) best. They are also clearly better than previous spatial reconstructions in terms of e.g. the degree of openness and tree cover in the northernmost parts of Europe and the western coasts of Norway. Our future goal is to use these LCC reconstructions in climate modelling studies and to gain insight into the effect of past anthropogenic deforestation. It is outside the scope of this paper to provide a discussion of the LCC reconstructions in terms of historical changes in vegetation abundance, land cover and human impact over the past 6 000 years. However, the methods developed here provide (some of) the tools needed for such a discussion.

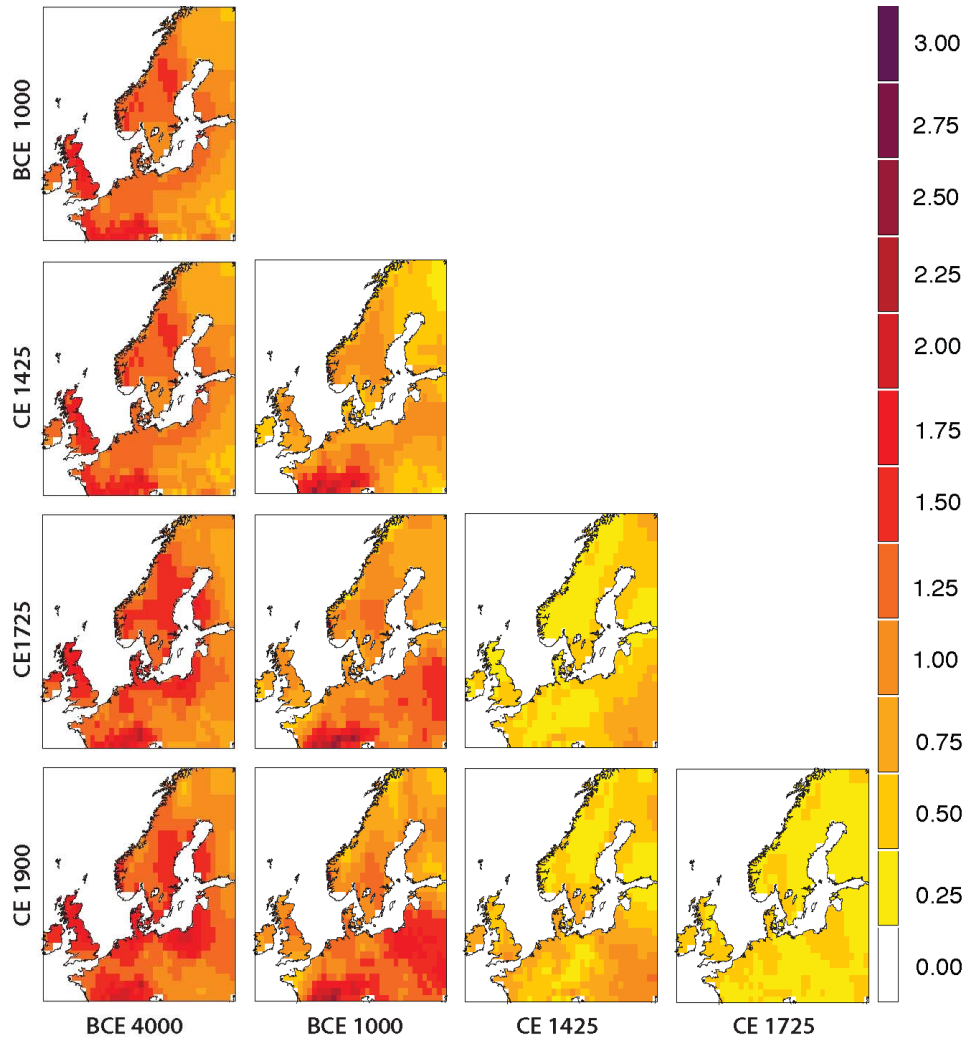


Figure 7: Compositional distances between the Full model LCC reconstructions for the different time periods.

## References

- J. Aitchison. *The statistical analysis of compositional data*. Chapman & Hall, Ltd., 1986.
- J. Aitchison, C. Barceló-Vidal, J. Martín-Fernández, and V. Pawłowsky-Glahn. Logratio analysis and compositional distance. *Math. Geol.*, 32(3):271–275, 2000.
- C. Andrieu and J. Thoms. A tutorial on adaptive mcmc. *Statist. and Comput.*, 18(4):343–373, 2008.
- J. J. Becker, D. T. Sandwell, W. H. F. Smith, J. Braud, B. Binder, J. Depner, D. Fabre, J. Factor, S. Ingalls, S. H. Kim, R. Ladner, K. Marks, S. Nelson, A. Pharaoh, G. Sharman, R. Trimmer, J. VonRosenburg, G. Wallace, and P. Weatherall. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30-PLUS. *Mar. Geod.*, 32(4):355–371, 2009.
- D. Billheimer, P. Guttorp, and W. F. Fagan. Statistical interpretation of species composition. *J. Am. Statist. Assoc.*, 96(456):1205–1214, 2001.
- V. Brovkin, M. Claussen, E. Driesschaert, T. Fichefet, D. Kicklighter, M. Loutre, H. Matthews, N. Ramankutty, M. Schaeffer, and A. Sokolov. Biogeophysical effects of historical land cover changes simulated by six earth system models of intermediate complexity. *Clim. Dynam.*, 26(6):587–600, 2006.
- M. Claussen, V. Brovkin, and A. Ganopolski. Biogeophysical versus biogeochemical feedbacks of large-scale land cover change. *Geophys. Res. Lett.*, 28(6):1011–1014, 2001.
- J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.
- M.-J. Gaillard, S. Sugita, F. Mazier, A.-K. Trondman, A. Brostrom, T. Hickler, J. O. Kaplan, E. Kjellström, U. Kokfelt, P. Kuneš, , C. Lemmen, P. Miller, J. Olofsson, A. Poska, M. Rundgren, B. Smith, G. Strandberg, R. Fyfe, A. Nielsen, T. Alenius, L. Balakauskas, L. Barnekov, H. Birks, A. Bjune, L. Björkman, T. Giesecke, K. Hjelle, L. Kalnina, M. Kangur, W. van der Knaap, T. Koff, P. Lagerås, M. Latałowa, M. Leydet, J. Lechterbeck, M. Lindbladh, B. Odgaard, S. Peglar, U. Segerström, H. von Stedingk, and H. Seppä. Holocene land-cover reconstructions for studies on land cover-climate feedbacks. *Clim. Past.*, 6:483–499, 2010.
- M.-J. Gaillard, T. Kleinen, P. Samuelsson, A.-B. Nielsen, J. Bergh, J. Kaplan, A. Poska, C. Sandström, G. Strandberg, A.-K. Trondman, and A. Wramneby. Causes of regional changeland cover. In *Second Assessment of Climate Change for the Baltic Sea Basin*, Regional Climate Studies, pages 453–477. Springer International Publishing, 2015.
- M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *J. Roy. Statist. Soc. Ser. B*, 73(2):123–214, 2011.
- G. H. Givens and J. A. Hoeting. *Computational statistics*, volume 710. John Wiley & Sons, 2012.

- S. Hellman, M.-j. Gaillard, A. Broström, and S. Sugita. Effects of the sampling design and selection of parameter values on pollen-based quantitative reconstructions of regional vegetation: a case study in southern Sweden using the reveals model. *Veg. Hist. Archaeobot.*, 17(5):445–459, 2008.
- J. O. Kaplan, K. M. Krumhardt, and N. Zimmermann. The prehistoric and preindustrial deforestation of Europe. *Quaternary. Sci. Rev.*, 28(27):3016–3034, 2009.
- K. Klein Goldewijk, A. Beusen, G. Van Dreht, and M. De Vos. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global. Ecol. Biogeogr.*, 20(1):73–86, 2011.
- L. Knorr-Held and H. Rue. On block updating in Markov random field models for disease mapping. *Scand. J. Statist.*, 29(4):597–614, 2002.
- F. Lindgren, R. Håvard, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. Roy. Statist. Soc. Ser. B*, 73(4):423–498, 2011.
- C. J. Paciorek and J. S. McLachlan. Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. *J. Am. Statist. Assoc.*, 104(486):608–622, 2009.
- R. Päivinen, M. Lehtikoinen, A. Schuck, T. Häme, S. Väättäinen, P. Kennedy, and S. Förling. *Combining earth observation data and forest statistics*. EuroForIns, 2001.
- B. Pirzamanbein, J. Lindström, A. Poska, S. Sugita, A.-K. Trondman, R. Fyfe, F. Mazier, A. B. Nielsen, J. O. Kaplan, A. E. Björne, H. J. B. Birks, T. Giesecke, M. Kangur, M. Latałowa, L. Marquer, B. Smith, and M.-J. Gaillard. Creating spatially continuous maps of past land cover from point estimates: A new statistical approach applied to pollen data. *Ecol. Complex.*, 20(0):127 – 141, 2014.
- J. Pongratz, C. Reick, T. Raddatz, and M. Claussen. A reconstruction of global agricultural areas and land cover for the last millennium. *Global. Biogeochem. Cy.*, 22(3), 2008.
- G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Statist. Soc. Ser. B*, 60(1):255–268, 1998.
- H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2004.
- A. Schuck, J. van Brusselen, R. Päivinen, T. Häme, P. Kennedy, and S. Förling. Compilation of a calibrated European forest map derived from NOAA-AVHRR data. EFI Internal Report 13, EuroForIns, 2002.
- B. Smith, I. C. Prentice, and M. T. Sykes. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space. *Global. Ecol. Biogeogr.*, 10(6): 621–637, 2001.

- G. Strandberg, E. Kjellström, A. Poska, S. Wagner, M.-J. Gaillard, A.-K. Trondman, A. Mauri, B. A. S. Davis, J. O. Kaplan, H. J. B. Birks, A. E. Bjune, R. Fyfe, T. Giesecke, L. Kalnina, M. Kangur, W. O. van der Knaap, U. Kokfelt, P. Kuneš, M. Latałowa, L. Marquer, F. Mazier, A. B. Nielsen, B. Smith, H. Seppä, and S. Sugita. Regional climate model simulations for europe at 6 and 0.2 k bp: sensitivity to changes in anthropogenic deforestation. *Clim. Past.*, 10(2):661–680, 2014.
- S. Sugita. Theory of quantitative reconstruction of vegetation II: all you need is love. *The Holocene*, 17(2):243–257, 2007a.
- S. Sugita. Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. *The Holocene*, 17(2):229–241, 2007b.
- H. Tjelmeland and K. V. Lund. Bayesian modelling of spatial compositional data. *J. Appl. Stat.*, 30(1):87–100, 2003.
- A.-K. Trondman, M.-J. Gaillard, F. Mazier, S. Sugita, R. Fyfe, A. B. Nielsen, C. Twiddle, P. Barratt, H. J. B. Birks, A. E. Bjune, L. Björkman, A. Broström, C. Caseldine, R. David, J. Dodson, W. Dörfler, E. Fischer, B. van Geel, T. Giesecke, T. Hultberg, L. Kalnina, M. Kangur, P. van der Knaap, T. Koff, P. Kuneš, P. Lagerås, M. Latałowa, J. Lechterbeck, C. Leroyer, M. Leydet, M. Lindbladh, L. Marquer, F. J. G. Mitchell, B. V. Odgaard, S. M. Peglar, T. Persson, A. Poska, M. Rösch, H. Seppä, S. Veski, and L. Wick. Pollen-based quantitative reconstructions of holocene regional vegetation cover (plant-functional types and land-cover types) in europe suitable for climate modelling. *Glob. Change Biol.*, 21(2):676–697, 2015.

## A Derivatives and Fisher Information of $[\boldsymbol{\eta}, \alpha | \mathbf{Y}]$

To construct the MALA updates for  $\boldsymbol{\eta}, \alpha$  we need the derivatives and Fisher information of the log-posterior (7),

$$\begin{aligned}
l(\boldsymbol{\eta}, \alpha | \mathbf{Y}) &= \log \left( \prod_{s=1}^{N_o} \mathbb{P}(\mathbf{y}_s | f(\boldsymbol{\eta}), \alpha, \mathbf{w}) \mathbb{P}(\mathbf{X} | \boldsymbol{\kappa}, \boldsymbol{\rho}) \mathbb{P}(\boldsymbol{\beta} | q_{\boldsymbol{\beta}}) \mathbb{P}(\alpha) \right) \\
&= \sum_{s=1}^{N_o} \log \Gamma(\alpha w_s) - \sum_{s=1}^{N_o} \sum_{k=1}^D \log \Gamma(\alpha w_s z_{s,k}) \\
&\quad + \sum_{s=1}^{N_o} \sum_{k=1}^D (\alpha w_s z_{s,k} - 1) \log y_{s,k} - \frac{1}{2} \mathbf{X}^\top (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\boldsymbol{\kappa})) \mathbf{X} \\
&\quad - \frac{q_{\boldsymbol{\beta}}}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} + (a_\alpha - 1) \log(\alpha) - \alpha b_\alpha + \text{const.}
\end{aligned} \tag{16}$$

Here  $\boldsymbol{\eta}$  is the latent  $\mathbb{R}^d$ -field at observed locations,  $\{u_s\}_{s=1}^{N_o}$ ,  $\mathbf{z}_s$  is the corresponding  $D$ -composition (i.e. defined on  $(0, 1)^D$ , with  $d = D - 1$ ), and const is an additive constant. Before computing derivatives of the log-posterior,  $l(\boldsymbol{\eta}, \alpha | \mathbf{Y})$ , we need some results for the compositional transformation.

### A.1 Derivatives of Compositional Transforms

The compositional transform used in this paper is the additive log-ratio (alr), (3), with inverse

$$z_k = \begin{cases} \frac{\exp(\eta_k)}{1 + \sum_{k=1}^d \exp(\eta_k)}, & \text{if } k = 1, \dots, D-1 \\ \frac{1}{1 + \sum_{k=1}^d \exp(\eta_k)}, & \text{if } k = D. \end{cases} \tag{17}$$

Here  $z$  is a  $D$ -compositional value (i.e.  $(0, 1)^D$ ) and  $\eta$  is  $\mathbb{R}^{D-1}$ .

For the MALA-computations the first and second derivatives of the inverse transformation are needed. These can be expressed in terms of the compositions,  $z_k$ ; for the first derivatives

$$\frac{\partial z_k}{\partial \eta_i} = \begin{cases} z_k(1 - z_k) & \text{if } k = i, \\ -z_k z_i & \text{if } k \neq i. \end{cases} \tag{18}$$

and for the second derivatives

$$\frac{\partial^2 z_k}{\partial \eta_i \partial \eta_j} = \begin{cases} z_k(1 - z_k)(1 - 2z_k), & \text{if } i = j, k = i \\ -z_k z_i(1 - 2z_i), & \text{if } i = j, k \neq i \\ -z_j z_k(1 - 2z_k), & \text{if } i \neq j, k = i \\ 2z_k z_i z_j, & \text{if } i \neq j, k \neq i, k \neq j, \end{cases} \tag{19}$$

the case  $i \neq j, k = j$  is obtained by symmetry.

One consequence of the sum to one constraint of compositional data is that

the derivatives (18) and second derivatives (19) also sum to one:

$$\begin{aligned}
\sum_{k=1}^D \frac{\partial z_k}{\partial \eta_i} &= z_i(1 - z_i) - \sum_{k \neq i} z_i z_k = z_i \left( 1 - \sum_{k=1}^D z_k \right) = 0 \\
\sum_{k=1}^D \frac{\partial^2 z_k}{\partial \eta_i^2} &= z_i(1 - z_i)(1 - 2z_i) - \sum_{k \neq i} z_k z_i(1 - 2z_i) = 0 \\
\sum_{k=1}^D \frac{\partial^2 z_k}{\partial \eta_i \partial \eta_j} &= -z_j z_i(1 - 2z_i) - z_i z_j(1 - 2z_j) + \sum_{k \neq i, j} 2z_k z_i z_j = 0
\end{aligned} \tag{20}$$

## A.2 Derivative of $l(\boldsymbol{\eta}, \alpha | \mathbf{Y})$

Recall that the latent field  $\boldsymbol{\eta}$  is a linear combination of the mean zero spatial field(s)  $\mathbf{X}$  and the regression coefficients  $\boldsymbol{\beta}$  given as

$$\boldsymbol{\eta} = \mathbf{A} \begin{pmatrix} \mathbb{I} & \mathbf{B} \end{pmatrix} \begin{bmatrix} \mathbf{X} \\ \boldsymbol{\beta} \end{bmatrix}.$$

Therefore, the updates of  $\boldsymbol{\eta}$  are done by updating the underlying fields and regression coefficients. Thus we need the derivatives of  $l(\boldsymbol{\eta}, \alpha | \mathbf{Y})$ , i.e.  $\nabla l$ , w.r.t  $\boldsymbol{\theta} = [\mathbf{X}^\top \quad \boldsymbol{\beta}^\top]^\top$  and  $\alpha$ .

$$\nabla_{\boldsymbol{\theta}} l(\boldsymbol{\eta}, \alpha | \mathbf{Y}) = [\mathbf{A} \quad \mathbf{AB}]^\top \nabla_{\boldsymbol{\eta}} l(\boldsymbol{\eta}, \alpha | \mathbf{Y}) - \begin{bmatrix} (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa)) \mathbf{X} \\ q_{\boldsymbol{\beta}} \boldsymbol{\beta} \end{bmatrix} \tag{21a}$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\eta}, \alpha | \mathbf{y})}{\partial \alpha} &= \sum_{s=1}^N w_s \psi(\alpha w_s) - \sum_{s=1}^N \sum_{k=1}^D w_s z_{s,k} \psi(\alpha w_s z_{s,k}) \\
&\quad + \sum_{s=1}^N \sum_{k=1}^D w_s z_{s,k} \log y_{s,k} + \frac{a_\alpha - 1}{\alpha} - b_\alpha
\end{aligned} \tag{21b}$$

where  $\nabla_{\boldsymbol{\theta}} l$  is the gradient w.r.t.  $\boldsymbol{\theta}$  (a  $Nd$ -column vector) and  $\psi(\cdot)$  is the digamma function. The elements of the gradient  $\nabla_{\boldsymbol{\eta}} l(\boldsymbol{\eta}, \alpha | \mathbf{Y})$  (a  $N_o d$ -column vector) are

$$\frac{\partial l(\boldsymbol{\eta}, \alpha | \mathbf{Y})}{\partial \eta_{s,k}} = \sum_{l=1}^D \left( -\alpha w_s \psi(\alpha w_s z_{s,l}) + \alpha w_s \log y_{s,l} \right) \frac{\partial z_{s,l}}{\partial \eta_{s,k}},$$

where the derivatives,  $\partial z_{s,l} / \partial \eta_{s,k}$ , depend on the choice of link function (see (18) for the alr-case).

## A.3 The Fisher Information

The Fisher information used in the MALA updates is computed as the expectation of the Hessian over observations,  $\mathbf{Y}$ , given all parameters and latent fields:

$$\mathbf{FI} = -\mathbb{E}_{\mathbf{Y}}(\mathbf{H}(l) | \boldsymbol{\eta}, \alpha, \boldsymbol{\rho}, \kappa) = \begin{bmatrix} \mathbf{FI}_{\boldsymbol{\theta}, \boldsymbol{\theta}} & \mathbf{FI}_{\boldsymbol{\theta}, \alpha} \\ \mathbf{FI}_{\alpha, \boldsymbol{\theta}} & \mathbf{FI}_{\alpha, \alpha} \end{bmatrix} \tag{22}$$

where  $\mathbf{H}(l)$  is Hessian of  $l(\boldsymbol{\eta}, \alpha | \mathbf{Y})$ . The resulting matrix consists of four blocks: two with second derivatives w.r.t.  $\boldsymbol{\theta}$  and  $\alpha$ , and two with cross partial derivatives;

each of the blocks is described below. For brevity we use  $\mathbb{E}(\mathbf{H}(l) | \bullet)$  to denote the conditional expectation in (22), and note that

$$\mathbb{E}(\log y_{s,k} | \bullet) = \psi(\alpha w_s z_{s,k}) - \psi\left(\sum_{l=1}^D \alpha w_s z_{s,l}\right) = \psi(\alpha w_s z_{s,k}) - \psi(\alpha w_s). \quad (23)$$

Similar to (21a) the top left block can be written as

$$\mathbf{F}\mathbf{I}_{\theta,\theta} = [\mathbf{A} \quad \mathbf{AB}]^\top \mathbf{H}_\eta [\mathbf{A} \quad \mathbf{AB}] + \begin{bmatrix} \boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa) & \mathbf{0} \\ \mathbf{0} & q_\beta \mathbb{I} \end{bmatrix},$$

where  $\mathbf{H}_\eta$  is a symmetric  $N_o d \times N_o d$  matrix with elements

$$\mathbf{H}_\eta^{(sk,s'k')} = -\mathbb{E}\left(\frac{\partial^2 l(\boldsymbol{\eta}, \alpha | \mathbf{y})}{\partial \eta_{s,k} \partial \eta_{s',k'}} \middle| \bullet\right).$$

The elements in  $\mathbf{H}_\eta$  are indexed by their spatial location,  $s = 1, \dots, N_o$ , and which latent field,  $k = 1, \dots, d$ , they belong to (i.e. which transformed compositional component). For elements at different locations

$$\mathbf{H}_\eta^{(sk,s'k')} = 0, \quad \text{if } s \neq s',$$

leaving only

$$\begin{aligned} \mathbf{H}_\eta^{(sk,sk')} &= -\mathbb{E}\left(\frac{\partial^2 l(\boldsymbol{\eta}, \alpha | \mathbf{y})}{\partial \eta_{s,k} \partial \eta_{s,k'}} \middle| \bullet\right) \\ &= -\frac{\partial}{\partial \eta_{s,k'}} \mathbb{E}\left(\sum_{l=1}^D \left(-\alpha w_s \psi(\alpha w_s z_{s,l}) + \alpha w_s \log y_{s,l}\right) \frac{\partial z_{s,l}}{\partial \eta_{s,k}} \middle| \bullet\right) \\ &= \alpha^2 w_s^2 \sum_{l=1}^D \psi'(\alpha w_s z_{s,l}) \frac{\partial z_{s,l}}{\partial \eta_{s,k'}} \frac{\partial z_{s,l}}{\partial \eta_{s,k}} \\ &\quad + \alpha w_s \sum_{l=1}^D \left(\psi(\alpha w_s z_{s,l}) - \mathbb{E}(\log y_{s,l} | \bullet)\right) \frac{\partial^2 z_{s,l}}{\partial \eta_{s,k} \partial \eta_{s,k'}}. \end{aligned}$$

Using the expectations in (23) gives

$$\mathbf{H}_\eta^{(sk,sk')} = \alpha^2 w_s^2 \sum_{l=1}^D \psi'(\alpha w_s z_{s,l}) \frac{\partial z_{s,l}}{\partial \eta_{s,k'}} \frac{\partial z_{s,l}}{\partial \eta_{s,k}} + \alpha w_s \psi(\alpha w_s) \sum_{l=1}^D \frac{\partial^2 z_{s,l}}{\partial \eta_{s,k} \partial \eta_{s,k'}},$$

and with the sum to zero result in (20) the elements of  $\mathbf{H}_\eta$  simplify to

$$\mathbf{H}_\eta^{(sk,sk')} = \alpha^2 w_s^2 \sum_{l=1}^D \psi'(\alpha w_s z_{s,l}) \frac{\partial z_{s,l}}{\partial \eta_{s,k'}} \frac{\partial z_{s,l}}{\partial \eta_{s,k}}.$$

Derivation of (21a) w.r.t.  $\alpha$  gives

$$\mathbf{F}\mathbf{I}_{\theta,\alpha} = -[\mathbf{A} \quad \mathbf{AB}]^\top \mathbb{E}\left(\frac{\partial}{\partial \alpha} \nabla_{\boldsymbol{\eta}} l(\boldsymbol{\eta}, \alpha | \mathbf{y}) \middle| \bullet\right).$$



The part concerning the gradient,  $\nabla_{\boldsymbol{\eta}} l(\boldsymbol{\eta}, \alpha | \mathbf{y})$ , gives a column vector of length  $N_o d$  with elements

$$\begin{aligned} \mathbb{E} \left( \frac{\partial^2 l(\boldsymbol{\eta}, \alpha | \mathbf{y})}{\partial \alpha \partial \eta_{s,k}} \middle| \bullet \right) &= w_s \sum_{l=1}^D \left( -\psi(\alpha w_s z_{s,l}) - \alpha w_s z_{s,l} \psi'(\alpha w_s z_{s,l}) + \right. \\ &\quad \left. + \mathbb{E}(\log y_{s,l} | \bullet) \right) \frac{\partial z_{s,l}}{\partial \eta_{s,k}}, \\ &= -\alpha w_s^2 \sum_{l=1}^D z_{s,l} \psi'(\alpha w_s z_{s,l}) \frac{\partial z_{s,l}}{\partial \eta_{s,k}}. \end{aligned}$$

The last equality is obtained from (23) and (20). Symmetry gives that  $\mathbf{F} \mathbf{I}_{\boldsymbol{\theta}, \alpha} = \mathbf{F} \mathbf{I}_{\alpha, \boldsymbol{\theta}}^\top$ .

The last block of (22) is

$$\begin{aligned} \mathbf{F} \mathbf{I}_{\alpha, \alpha} &= -\mathbb{E} \left( \frac{\partial^2 l(\boldsymbol{\eta}, \alpha | \mathbf{y})}{\partial \alpha^2} \middle| \bullet \right) \\ &= - \left( \sum_{s=1}^N w_s^2 \psi'(\alpha w_s) - \sum_{s=1}^N \sum_{k=1}^D w_s^2 z_{s,k}^2 \psi'(\alpha w_s z_{s,k}) - \frac{a_\alpha - 1}{\alpha^2} \right). \end{aligned}$$

## B The Posterior $\kappa|\mathbf{X}$

The posterior of  $\kappa|\mathbf{X}$  is obtained by integrating out  $\boldsymbol{\rho}$  from the joint posterior of  $\kappa, \boldsymbol{\rho}|\mathbf{X}$ . With the densities for  $\mathbf{X}$  and  $\boldsymbol{\rho}$  given as

$$\mathbf{X}|\kappa, \boldsymbol{\rho} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{\rho} \otimes \mathbf{Q}(\kappa))^{-1}) \quad \text{and} \quad \boldsymbol{\rho} \sim IW(\mathbb{I}, df) \quad (24)$$

in (6) the posterior  $\kappa|\mathbf{X}$  is

$$\begin{aligned} \mathbb{P}(\kappa|\mathbf{X}) &\propto \int \mathbb{P}(\mathbf{X}|\kappa, \boldsymbol{\rho}) \mathbb{P}(\kappa) \mathbb{P}(\boldsymbol{\rho}) d\boldsymbol{\rho} \\ &\propto \int |\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa)|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{X}^\top (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa)) \mathbf{X}\right) \mathbb{P}(\kappa) \\ &\quad \cdot |\boldsymbol{\rho}|^{-\frac{df+d+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\rho}^{-1})\right) d\boldsymbol{\rho}. \end{aligned} \quad (25)$$

Introducing vectorization such that  $\text{vec}(\mathbf{x}) = \mathbf{X}$ , where  $\mathbf{x} = (X_1, \dots, X_d)$  is a  $N \times d$ -matrix version of the column-vector  $\mathbf{X}$ , the exponential term can be rewritten as

$$\begin{aligned} -\frac{1}{2} \mathbf{X}^\top (\boldsymbol{\rho}^{-1} \otimes \mathbf{Q}(\kappa)) \mathbf{X} &= -\frac{1}{2} \mathbf{X}^\top \text{vec}(\mathbf{Q}(\kappa)^\top \mathbf{x} \boldsymbol{\rho}^{-1}) \\ &= -\frac{1}{2} \text{tr}(\mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x} \boldsymbol{\rho}^{-1}) = -\frac{1}{2} \text{tr}(\boldsymbol{\rho}^{-1} \mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x}). \end{aligned}$$

The posterior in (25) now simplifies to

$$\mathbb{P}(\kappa|\mathbf{X}) \propto \mathbb{P}(\kappa) |\mathbf{Q}(\kappa)|^{\frac{d}{2}} \int |\boldsymbol{\rho}|^{-\frac{N+df+d+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\rho}^{-1} (\mathbb{I} + \mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x}))\right) d\boldsymbol{\rho}.$$

Recognizing the density of an unnormalized inverse-Wishart distribution under the integral sign we normalise and obtain the posteriors

$$\begin{aligned} \boldsymbol{\rho}|\kappa, \mathbf{X} &\sim IW(\mathbb{I} + \mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x}, df + N), \\ \mathbb{P}(\kappa|\mathbf{X}) &\propto \mathbb{P}(\kappa) \cdot \frac{|\mathbf{Q}(\kappa)|^{\frac{d}{2}}}{|\mathbb{I} + \mathbf{x}^\top \mathbf{Q}(\kappa) \mathbf{x}|^{\frac{N+df}{2}}}. \end{aligned}$$

## C Parameter Estimates

Table 5: Parameter estimates (Est) and 95% quantile (CI) for the two models used to reconstruct past LCC from PbLCC data.

1700 CE				
Parameter	Full		RM	
	Est	(CI)	Est	(CI)
$\alpha$	9.55	( 7.64 , 12.93 )	6.07	( 5.31 , 6.86 )
$\kappa$	0.23	( 0.12 , 0.39 )	-	-
$\rho_{11}$	0.50	( 0.13 , 1.83 )	-	-
$\rho_{12}$	0.23	( 0.01 , 1.11 )	-	-
$\rho_{22}$	0.25	( 0.07 , 0.90 )	-	-
$\beta_{10}$	-0.72	( -1.85 , 0.23 )	-0.13	( -0.24 , 0.00 )
$\beta_{11}$	0.17	( 0.08 , 0.26 )	0.28	( 0.25 , 0.31 )
$\beta_{12}$	-0.01	( -0.12 , 0.10 )	-0.08	( -0.12 , -0.03 )
$\beta_{13}$	-0.01	( -0.20 , 0.18 )	-0.14	( -0.24 , -0.05 )
$\beta_{20}$	-0.74	( -1.56 , 0.02 )	-0.35	( -0.49 , -0.23 )
$\beta_{21}$	0.07	( -0.01 , 0.15 )	0.13	( 0.11 , 0.16 )
$\beta_{22}$	-0.05	( -0.13 , 0.03 )	-0.08	( -0.12 , -0.04 )
$\beta_{23}$	-0.20	( -0.38 , -0.03 )	-0.30	( -0.40 , -0.20 )

Table 6: Parameter estimates (Est) and 95% quantile (CI) for the two models used to reconstruct past LCC from PbLCC data.

1400 CE				
Parameter	Full		RM	
	Est	(CI)	Est	(CI)
$\alpha$	8.75	( 7.22 , 10.76 )	5.18	( 4.57 , 5.83 )
$\kappa$	0.18	( 0.08 , 0.31 )	-	-
$\rho_{11}$	0.37	( 0.10 , 0.98 )	-	-
$\rho_{12}$	0.12	( -0.02 , 0.47 )	-	-
$\rho_{22}$	0.17	( 0.06 , 0.44 )	-	-
$\beta_{10}$	-0.70	( -2.34 , 0.77 )	-0.09	( -0.21 , 0.03 )
$\beta_{11}$	0.16	( 0.05 , 0.26 )	0.28	( 0.26 , 0.31 )
$\beta_{12}$	0.02	( -0.10 , 0.13 )	-0.07	( -0.12 , -0.02 )
$\beta_{13}$	0.09	( -0.10 , 0.28 )	-0.09	( -0.19 , 0.00 )
$\beta_{20}$	-0.56	( -1.78 , 0.51 )	-0.15	( -0.28 , -0.03 )
$\beta_{21}$	0.07	( -0.03 , 0.16 )	0.14	( 0.11 , 0.17 )
$\beta_{22}$	-0.03	( -0.12 , 0.06 )	-0.07	( -0.11 , -0.03 )
$\beta_{23}$	-0.18	( -0.36 , -0.01 )	-0.32	( -0.42 , -0.22 )

Table 7: Parameter estimates (Est) and 95% quantile (CI) for the two models used to reconstruct past LCC from PbLCC data.

1000 BCE				
Parameter	Full		RM	
	Est	(CI)	Est	(CI)
$\alpha$	7.02	( 5.89 , 8.37 )	4.42	( 3.91 , 4.96 )
$\kappa$	0.19	( 0.08 , 0.30 )	-	-
$\rho_{11}$	0.32	( 0.10 , 0.80 )	-	-
$\rho_{12}$	0.07	( -0.04 , 0.27 )	-	-
$\rho_{22}$	0.16	( 0.06 , 0.37 )	-	-
$\beta_{10}$	0.19	( -1.40 , 1.57 )	0.50	( 0.37 , 0.63 )
$\beta_{11}$	0.24	( 0.13 , 0.35 )	0.30	( 0.27 , 0.33 )
$\beta_{12}$	-0.03	( -0.17 , 0.10 )	0.05	( -0.01 , 0.11 )
$\beta_{13}$	0.15	( -0.06 , 0.37 )	-0.02	( -0.12 , 0.09 )
$\beta_{20}$	0.19	( -1.07 , 1.21 )	0.55	( 0.43 , 0.68 )
$\beta_{21}$	0.07	( -0.01 , 0.16 )	0.12	( 0.09 , 0.14 )
$\beta_{22}$	0.03	( -0.08 , 0.13 )	0.02	( -0.03 , 0.07 )
$\beta_{23}$	-0.02	( -0.20 , 0.18 )	-0.11	( -0.22 , -0.01 )

Table 8: Parameter estimates (Est) and 95% quantile (CI) for the two models used to reconstruct past LCC from PbLCC data.

4000 BCE				
Parameter	Full		RM	
	Est	(CI)	Est	(CI)
$\alpha$	7.58	( 6.26 , 9.84 )	5.36	( 4.72 , 6.02 )
$\kappa$	0.20	( 0.10 , 0.32 )	-	-
$\rho_{11}$	0.21	( 0.07 , 0.69 )	-	-
$\rho_{12}$	0.10	( -0.02 , 0.64 )	-	-
$\rho_{22}$	0.24	( 0.06 , 1.03 )	-	-
$\beta_{10}$	0.41	( -0.70 , 1.48 )	0.38	( 0.24 , 0.53 )
$\beta_{11}$	0.23	( 0.13 , 0.33 )	0.23	( 0.20 , 0.26 )
$\beta_{12}$	-0.19	( -0.36 , -0.03 )	-0.04	( -0.11 , 0.03 )
$\beta_{13}$	0.04	( -0.17 , 0.25 )	-0.04	( -0.14 , 0.07 )
$\beta_{20}$	0.61	( -0.66 , 1.58 )	0.99	( 0.85 , 1.12 )
$\beta_{21}$	0.01	( -0.09 , 0.10 )	0.08	( 0.06 , 0.11 )
$\beta_{22}$	-0.01	( -0.16 , 0.14 )	0.00	( -0.06 , 0.05 )
$\beta_{23}$	-0.05	( -0.24 , 0.15 )	-0.25	( -0.35 , -0.15 )

## D Maps of Estimated Land Cover

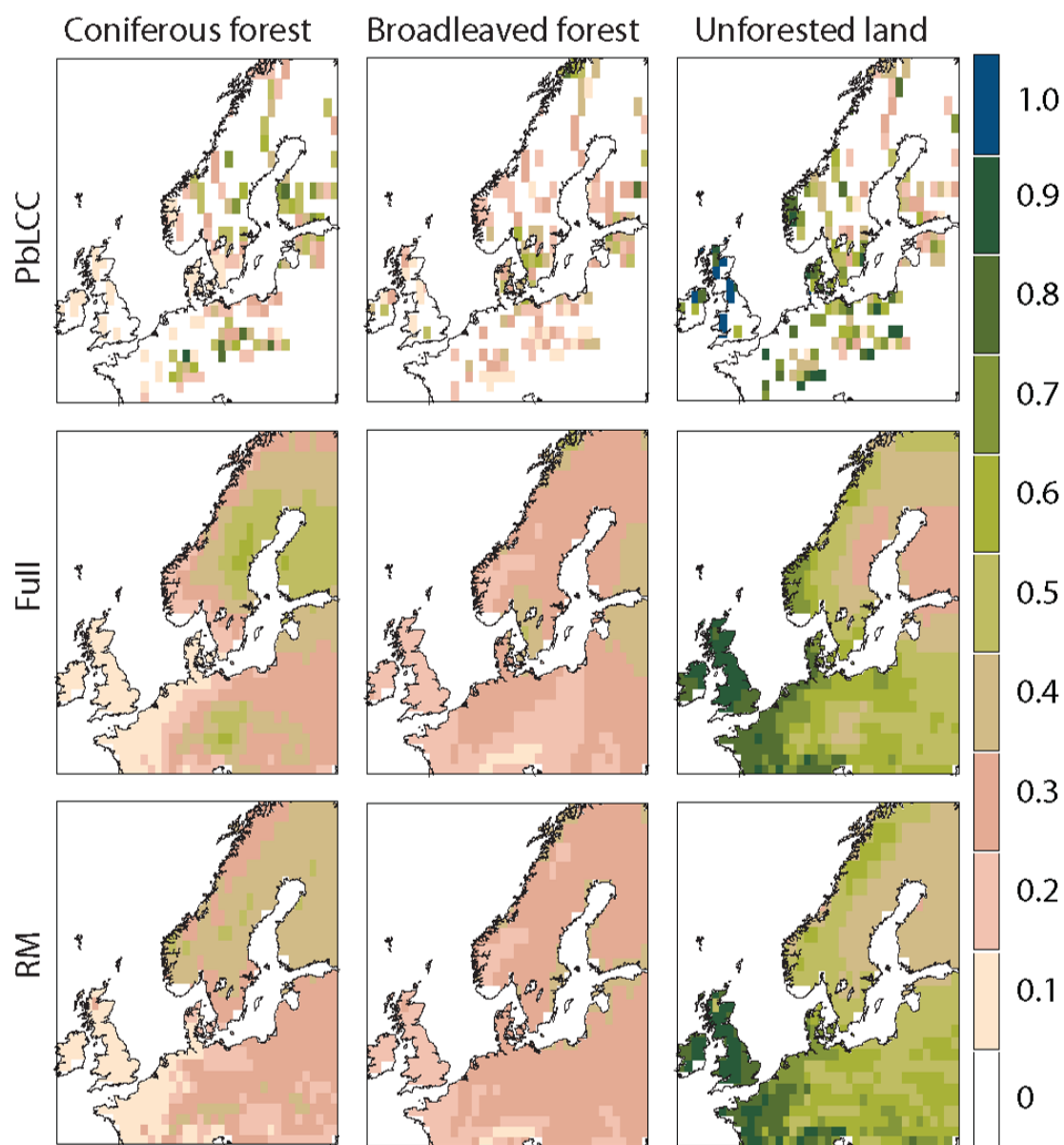


Figure 8: Results for the 1725 CE time period: the first row shows the PbLCC data, and the other rows show the reconstructions for the Full model and RM.

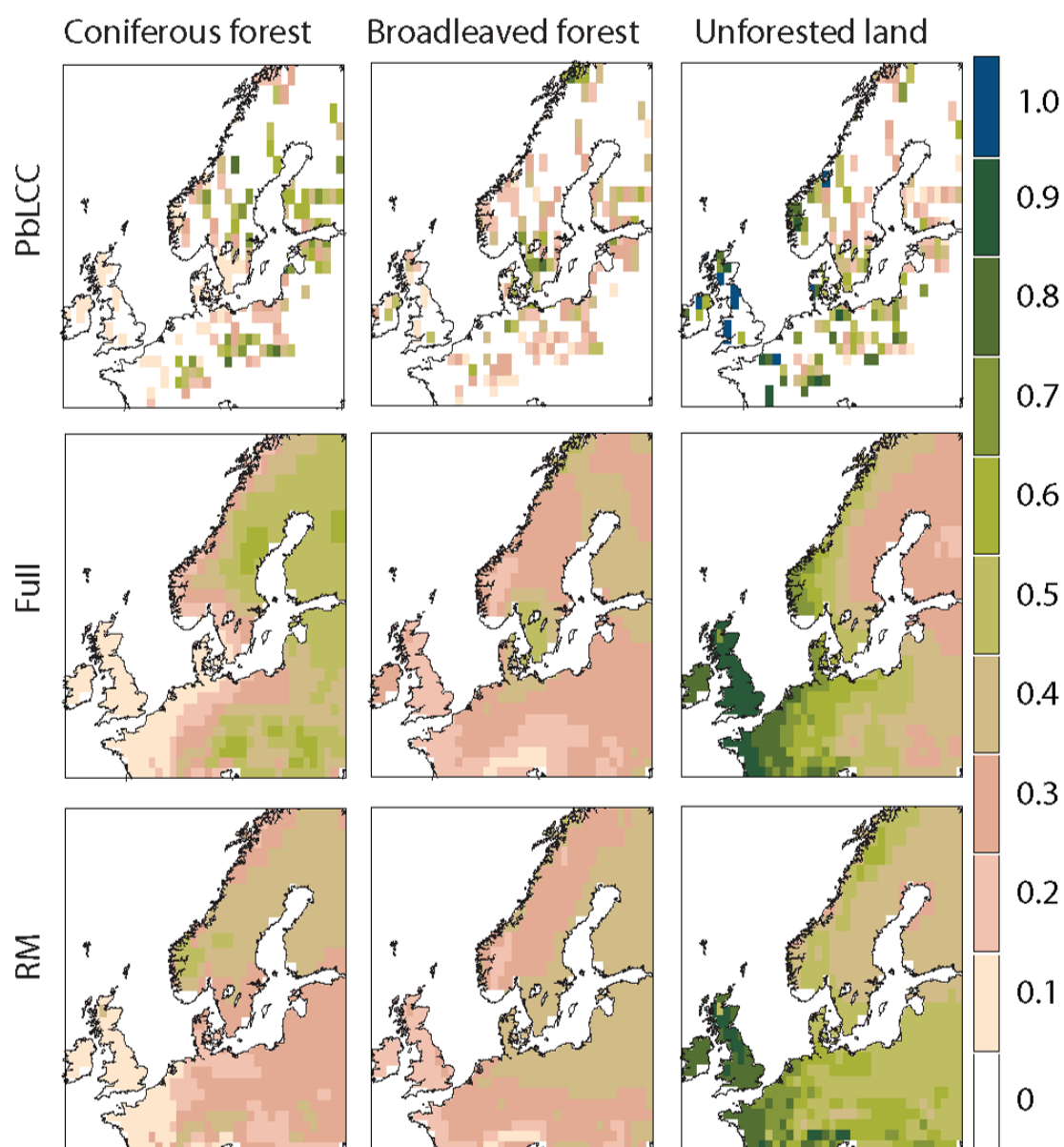


Figure 9: Results for the 1425 CE time period: the first row shows the PbLCC data, and the other rows show the reconstructions for the Full model and RM.

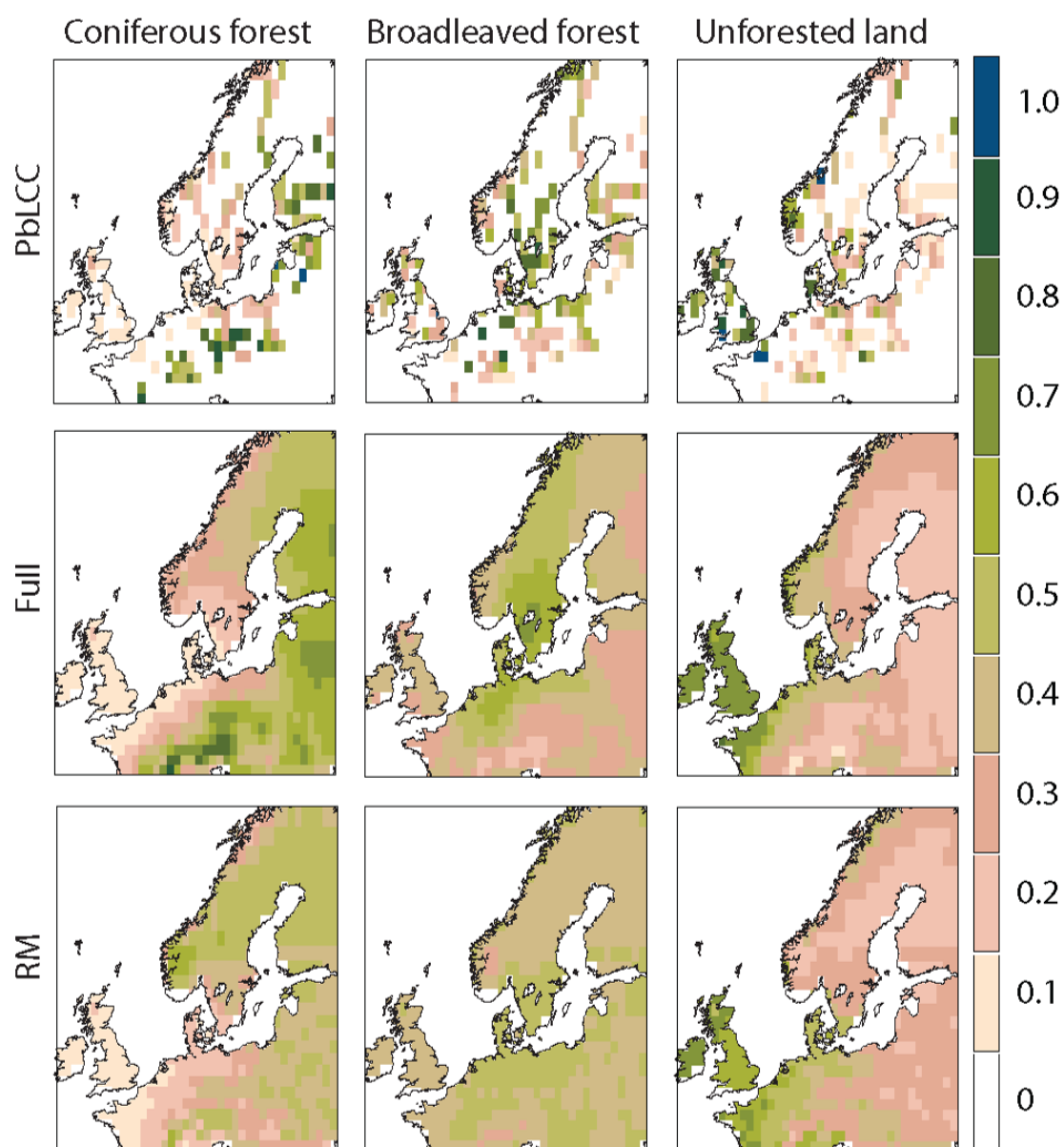


Figure 10: Results for the 1000 BCE time period: the first row shows the PbLCC data, and the other rows show the reconstructions for the Full model and RM.

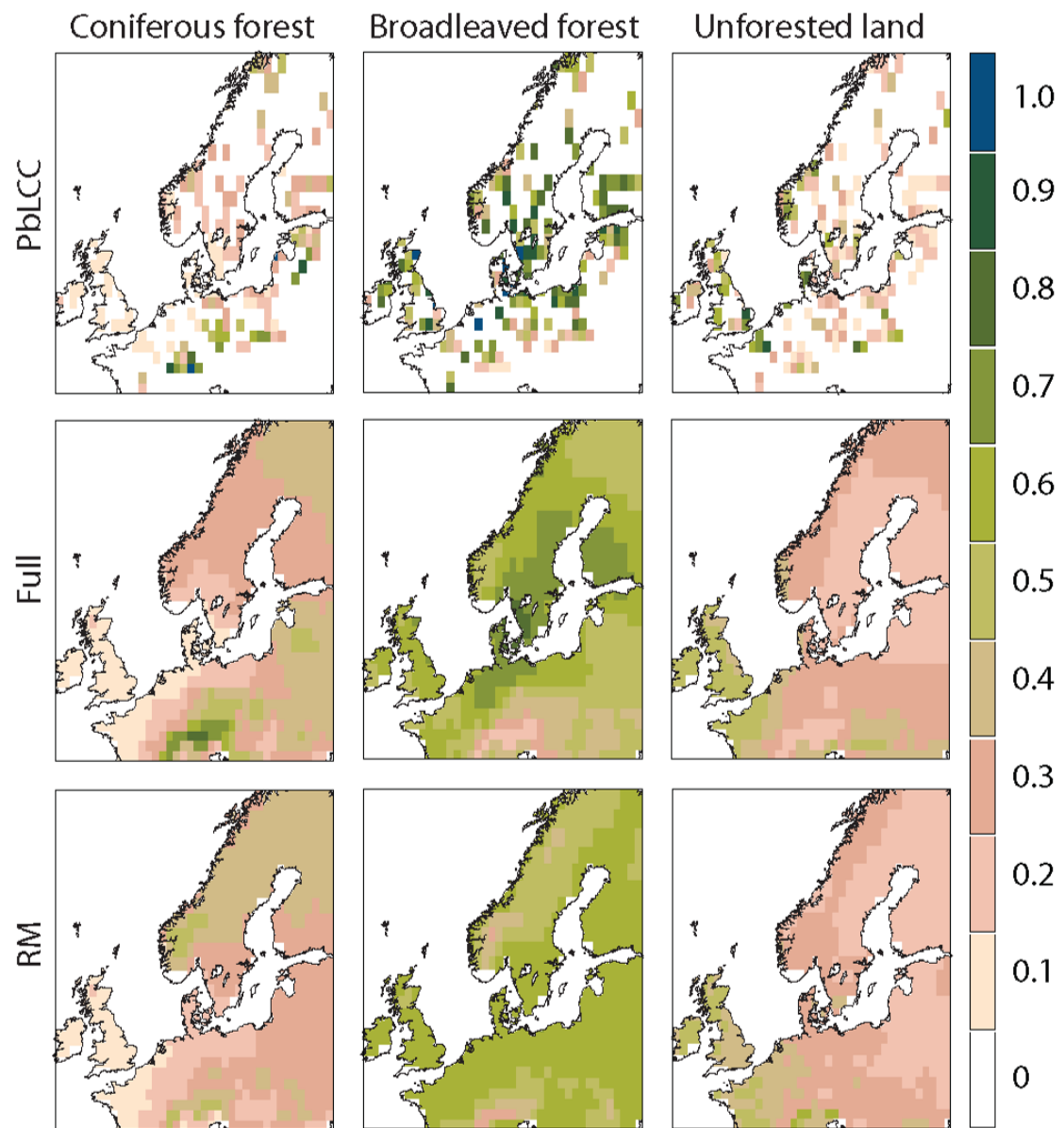


Figure 11: Results for the 4000 BCE time period: the first row shows the PbLCC data, and the other rows show the reconstructions for the Full model and RM.



## E Uncertainties in Estimated Land Cover

### E.1 Maps of Uncertainties

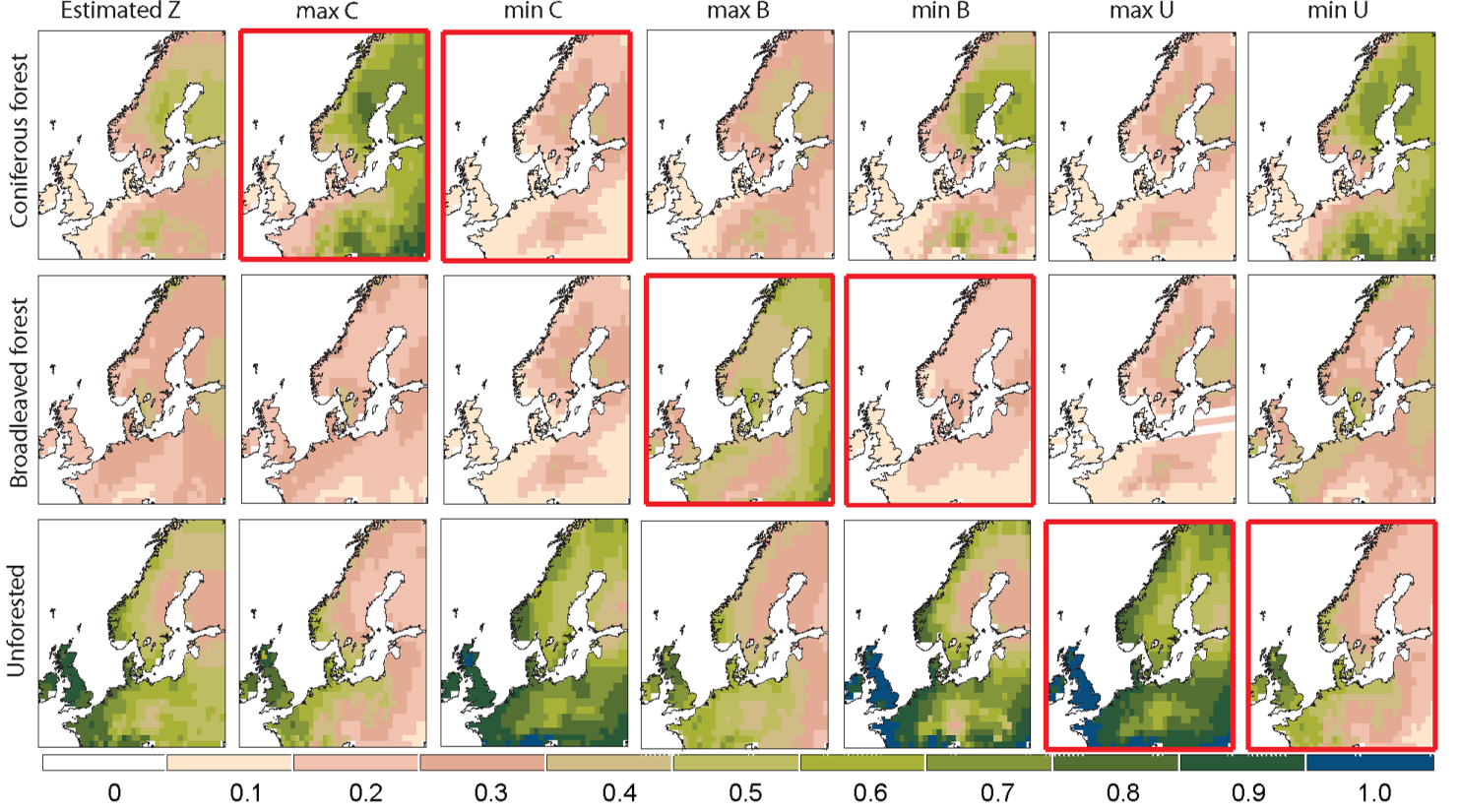


Figure 12: The first column shows the reconstructed LCC for the 1725 CE time period, using the Full model. Rows 2 and 3, column 1 (with tick axes), show the maximum and minimum of 95% elliptical CR (at each location) for Coniferous; columns 2 and 3 give the corresponding Broadleaved and Unforested compositions. Rows 4 and 5 (column 2 with tick axes) gives the bounds for the Broadleaved composition while rows 6 and 7 show the bounds for Unforested land (column 3 with tick axes). For each of the rows the corresponding changes in the remaining two compositions are also given.

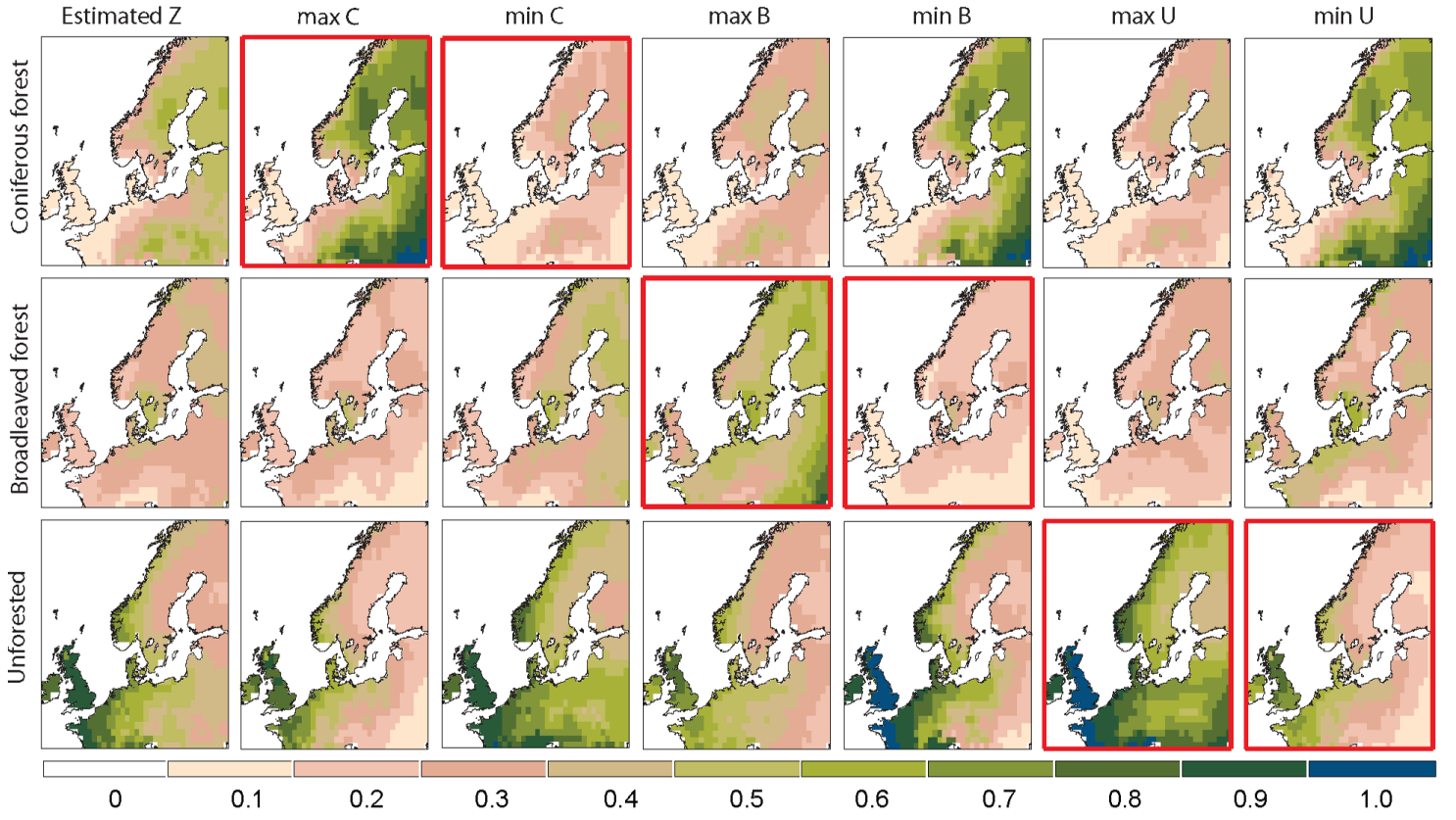


Figure 13: The first column shows the reconstructed LCC for the 1425 CE time period, using the Full model. Rows 2 and 3, column 1 (with tick axes), show the maximum and minimum of 95% elliptical CR (at each location) for Coniferous; columns 2 and 3 give the corresponding Broadleaved and Unforested compositions. Rows 4 and 5 (column 2 with tick axes) gives the bounds for the Broadleaved composition while rows 6 and 7 show the bounds for Unforested land (column 3 with tick axes). For each of the rows the corresponding changes in the remaining two compositions are also given.

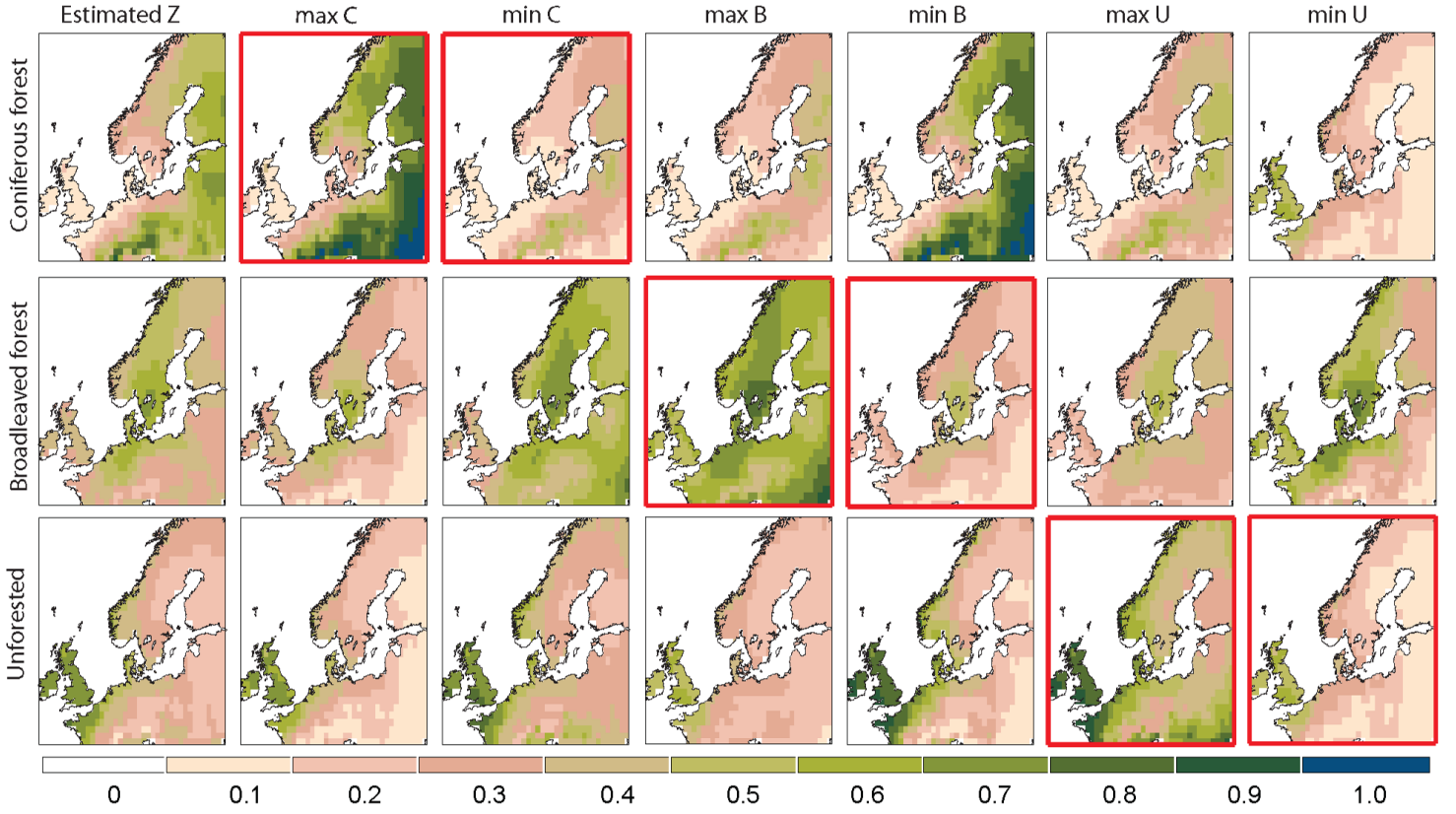


Figure 14: The first column shows the reconstructed LCC for the 1000 BCE time period, using the Full model. Rows 2 and 3, column 1 (with tick axes), show the maximum and minimum of 95% elliptical CR (at each location) for Coniferous; columns 2 and 3 give the corresponding Broadleaved and Unforested compositions. Rows 4 and 5 (column 2 with tick axes) gives the bounds for the Broadleaved composition while rows 6 and 7 show the bounds for Unforested land (column 3 with tick axes). For each of the rows the corresponding changes in the remaining two compositions are also given.

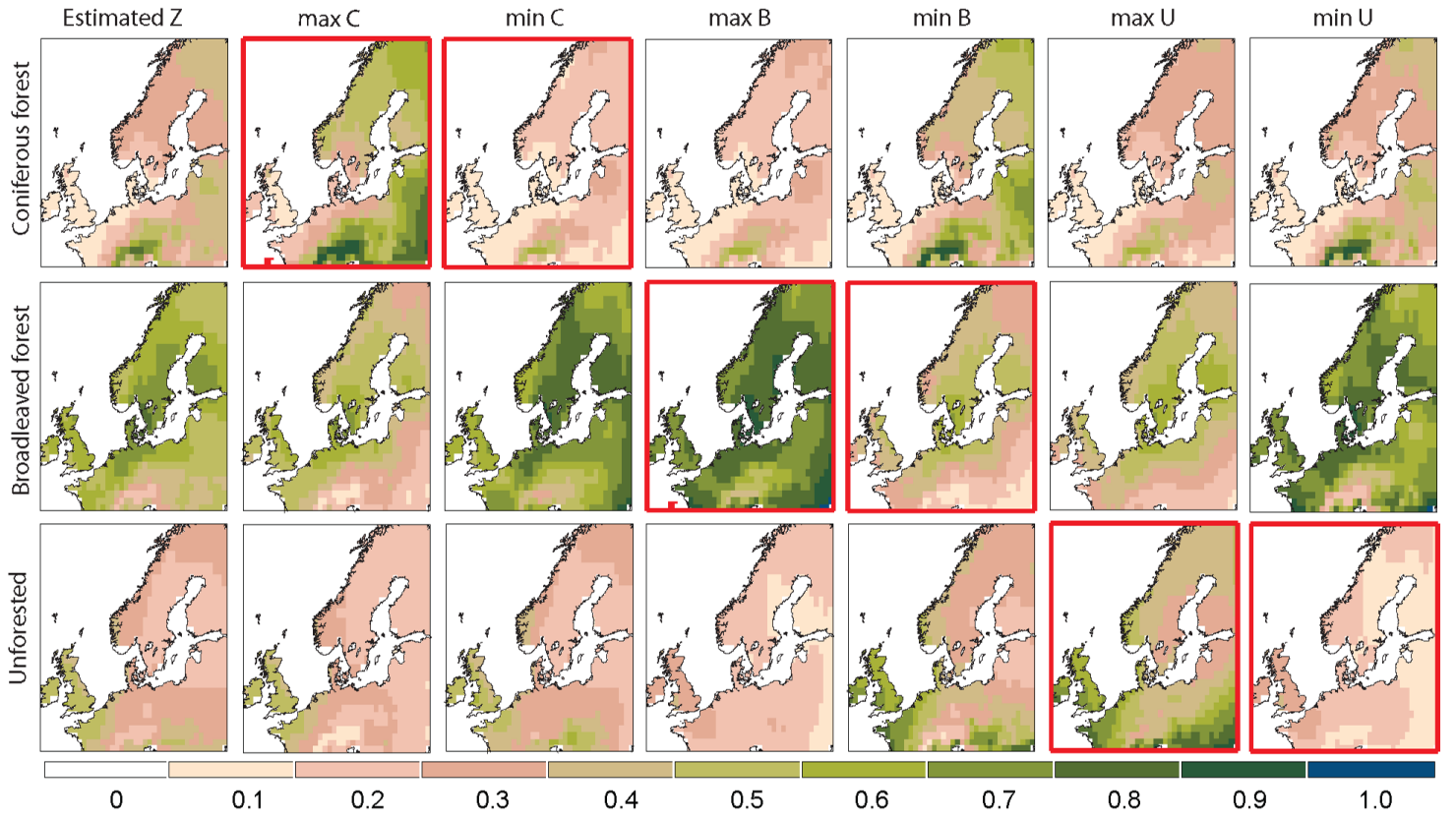


Figure 15: The first column shows the reconstructed LCC for the 4000 BCE time period, using the Full model. Rows 2 and 3, column 1 (with tick axes), show the maximum and minimum of 95% elliptical CR (at each location) for Coniferous; columns 2 and 3 give the corresponding Broadleaved and Unforested compositions. Rows 4 and 5 (column 2 with tick axes) gives the bounds for the Broadleaved composition while rows 6 and 7 show the bounds for Unforested land (column 3 with tick axes). For each of the rows the corresponding changes in the remaining two compositions are also given.

## E.2 Confidence Regions for Selected Locations

Confidence and Prediction region 1725 CE

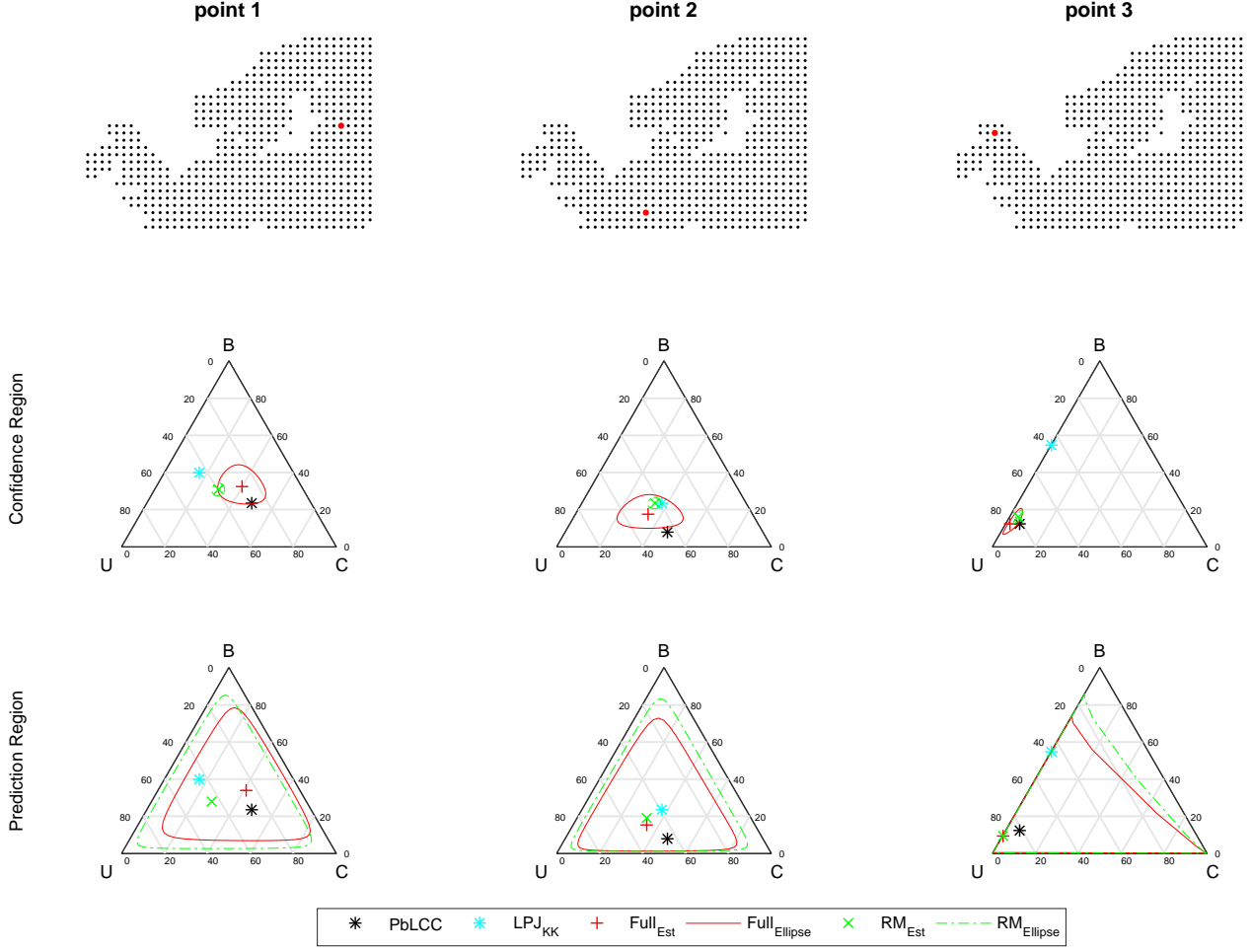


Figure 16: The first row shows the locations of the three selected grid cells for the 1725 CE time period. The second row shows the ternary CR along with the reconstructions for the two models and the values of the PbLCC data and the LPJ-GUESS<sub>KK</sub> land cover covariate at each location. The third row shows the ternary PRs and the same values.

## Confidence and Prediction region 1425 CE

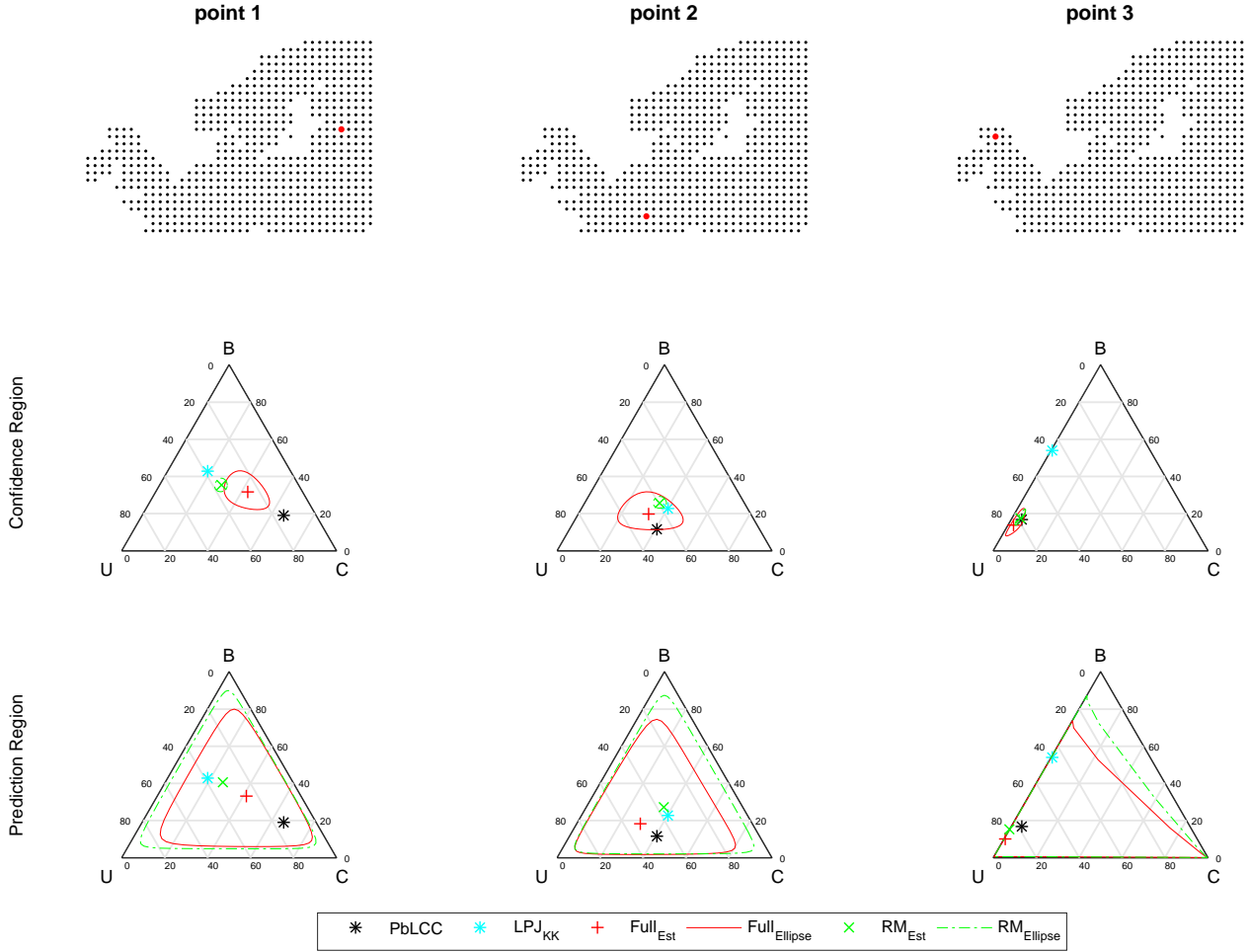


Figure 17: The first row shows the locations of the three selected grid cells for the 1425 CE time period. The second row shows the ternary CR along with the reconstructions for the two models and the values of the PbLCC data and the LPJ-GUESS<sub>KK</sub> land cover covariate at each location. The third row shows the ternary PRs and the same values.

## Confidence and Prediction region 1000 BCE

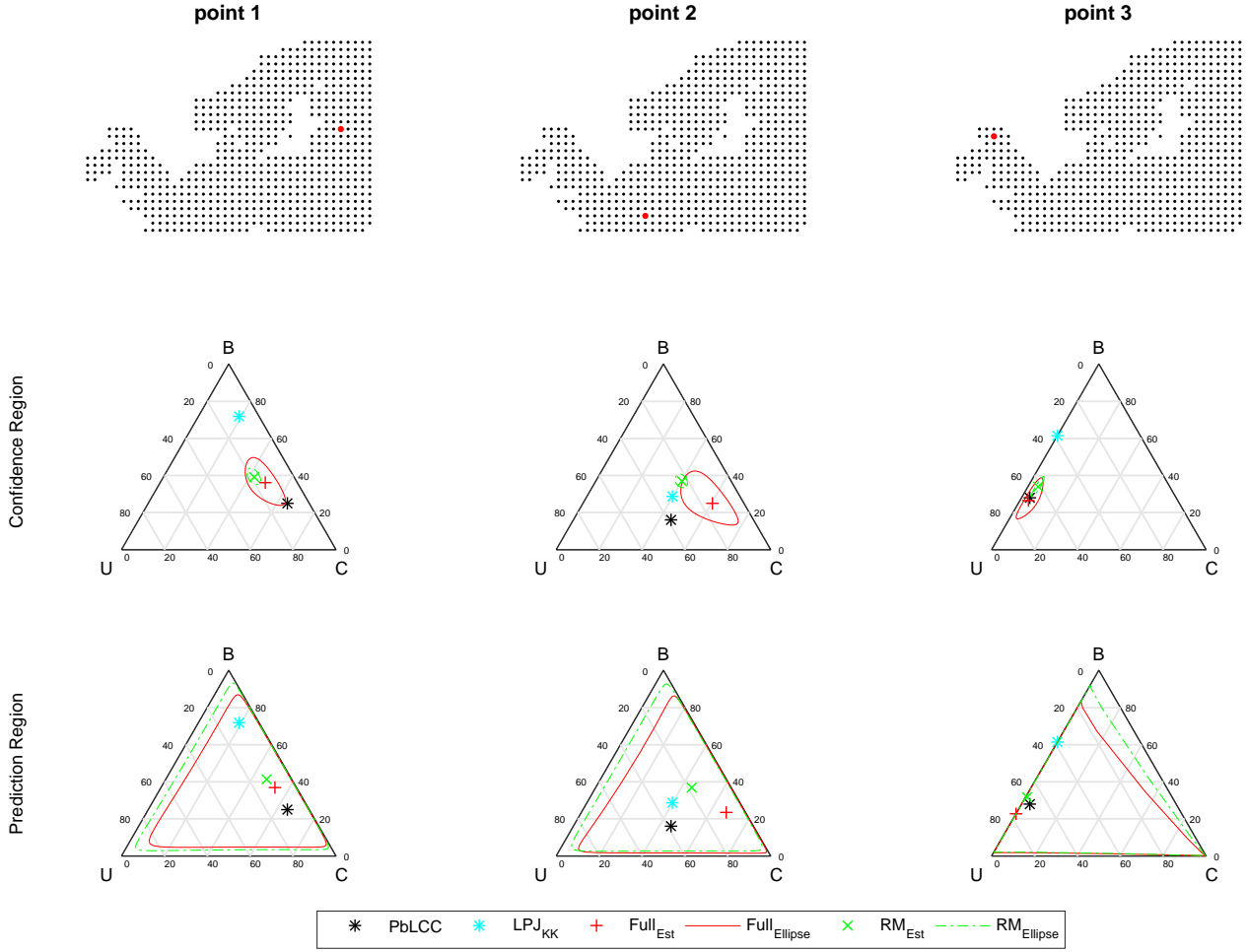


Figure 18: The first row shows the locations of the three selected grid cells for the 1000 BCE time period. The second row shows the ternary CR along with the reconstructions for the two models and the values of the PbLCC data and the LPJ-GUESS<sub>KK</sub> land cover covariate at each location. The third row shows the ternary PRs and the same values.

## Confidence and Prediction region 4000 BCE

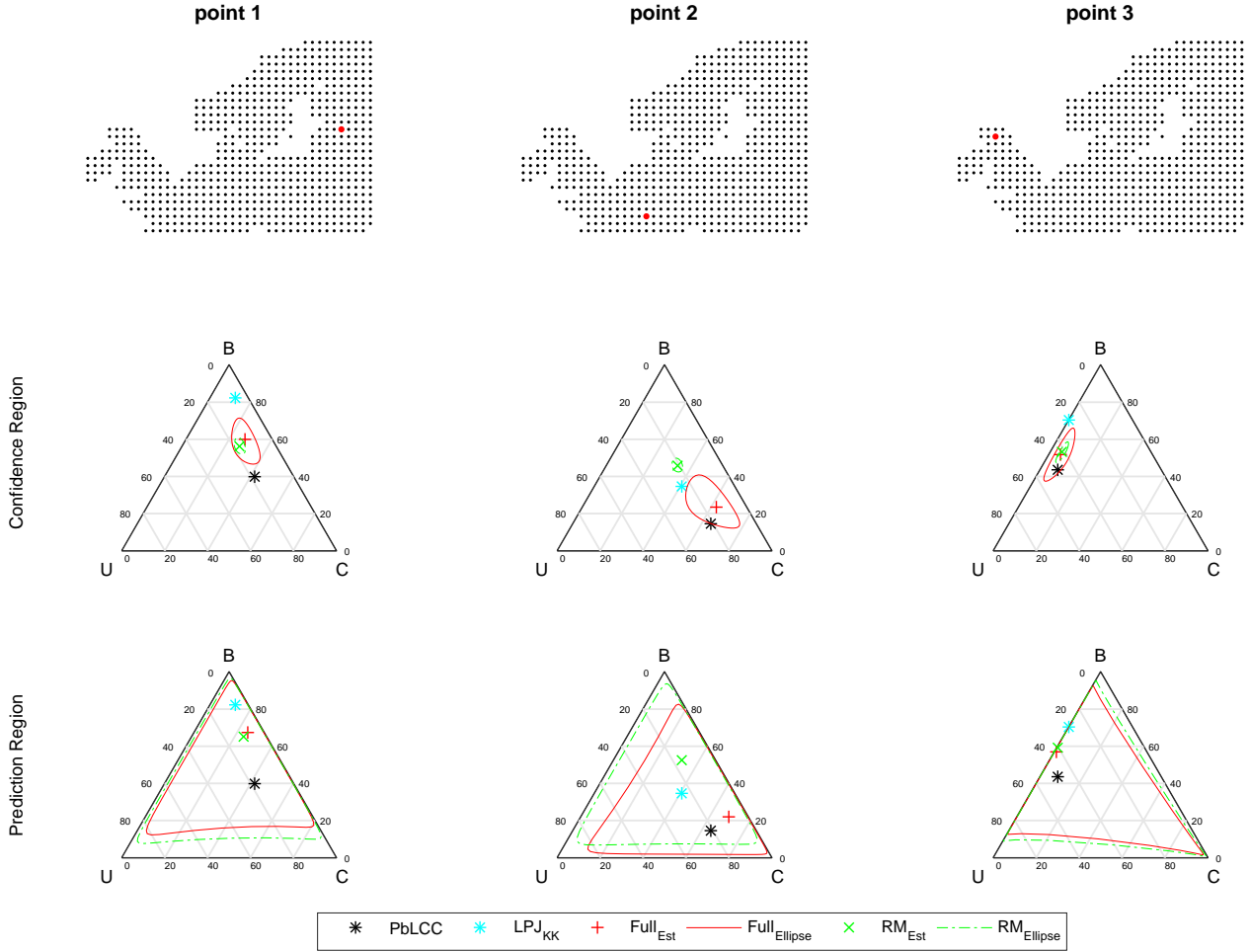


Figure 19: The first row shows the locations of the three selected grid cells for the 4000 BCE time period. The second row shows the ternary CR along with the reconstructions for the two models and the values of the PbLCC data and the LPJ-GUESS<sub>KK</sub> land cover covariate at each location. The third row shows the ternary PRs and the same values.